

AN EFFICIENT AND ROBUST TEST FOR CHANGE-POINTS IN CORRELATION

HEROLD DEHLING, DANIEL VOGEL, MARTIN WENDLER, AND DOMINIK WIED

ABSTRACT. For a bivariate time series $((X_i, Y_i))_{i=1, \dots, n}$ we want to detect whether the correlation between X_i and Y_i stays constant for all $i = 1, \dots, n$. We propose a nonparametric change-point test statistic based on Kendall's tau. The asymptotic distribution under the null hypothesis of no change follows from a new U -statistic invariance principle. Assuming a single change-point, we show that the location of the change-point is consistently estimated.

Kendall's tau possesses a high efficiency at the normal distribution, as compared to the normal maximum likelihood estimator, Pearson's moment correlation. Contrary to Pearson's correlation coefficient, it has excellent robustness properties and shows no loss in efficiency at heavy-tailed distributions. The motivation for this research article originates in financial data situations, where heavy tails are common and Kendall's tau is a more efficient estimator than the moment correlation.

We assume the data $((X_i, Y_i))_{i=1, \dots, n}$ to be stationary and P -near epoch dependent on an absolutely regular process. This large class of processes includes all common time series models as well as many chaotic dynamical systems. The P -near epoch dependence condition constitutes a generalization of the usually considered L_p -near epoch dependence, $p \geq 1$, that allows for arbitrarily heavy-tailed data.

We investigate the test numerically and compare it to previous proposals.

CONTENTS

1. Introduction	1
2. Statement of Main Results	2
3. Invariance Principles for P -NED Sequences	7
4. Change-point Identification	8
5. Properties of the Test and Comparison to Previous Proposals	9
6. Simulation Results	12
7. Data Examples	18
8. Conclusion	19
Acknowledgment	21
Appendix A. Proofs of Section 2	21
Appendix B. Proofs of Section 3	25
Appendix C. Proofs of Section 4	29
Appendix D. Misc	31
References	31

1. INTRODUCTION

The problem of detecting changes in the distribution of sequential observations has a long history in statistics, see e.g. Csörgő and Horváth (1997). However, particularly detecting

Key words and phrases. Change-point analysis, Kendall's tau, U -statistic, functional limit theorem, near epoch dependence.

changes in the dependence structure of multivariate time series has just recently attracted the focus of statistical research (Galeano and Peña, 2007; Aue, Hörmann, Horváth, and Reimherr, 2009; Kao, Trapani, and Urga, 2011).

The authors' interest in the field originates in the economic data analyses. For risk management and portfolio optimization the dependence between financial asset prices is of enormous importance. It is often assumed to be constant over the observed time period, which is a simplifying assumption that is evidently violated for longer observation periods. For good statistical modeling and successful decision making it is essential to detect changes in the association of financial price processes and, within reasonable time frames, re-estimate the correlation parameters. Particularly, in times of global financial crises, the price processes of most financial assets tend to be highly dependent, united in their common downward trend, causing the hedging powers of investment diversification to cease — an effect, for which the term *diversification meltdown* has been coined (Campbell et al., 2003).

Recently, Wied, Krämer, and Dehling (2012) proposed a test for change in correlation based on Pearson's moment correlation. We recommend to use Kendall's tau instead of Pearson's correlation coefficient because it possesses a higher efficiency at heavy tailed distributions. For details see Section 5.

The paper is organized as follows: Section 2 contains the main results, which concern the asymptotic behavior of the test statistic under the null hypothesis of no change in correlation. Theorem 2.6 gives the asymptotic distribution of the test statistic. The asymptotic distribution contains a long run variance parameter; Theorem 2.9 shows the consistency of a proposed estimator for the long run variance parameter and allows hence to formulate an asymptotically distribution-free version of the test statistic. Sections 3 contains the theoretical groundwork: two functional limit theorems for weakly dependent processes (Theorems 3.2 and 3.5), which are the cornerstones of the proofs of the results of Section 2, but which are also of interest in their own right. Section 4 deals with estimating the location of a potential change-point.

Section 6 examines the properties of the proposed statistical procedure numerically, and Section 7 contains applications to real life data examples. All proofs are deferred to the appendix.

We use bold type face to denote vector-valued objects. Throughout, $|\cdot|_p$ denotes the p -norm in \mathbb{R}^q , $p \in [1, \infty)$, $q \in \mathbb{N}$. To denote the L_p norm of a real-valued random variable X , we write $(E|X|^p)^{1/p}$, $p \in [1, \infty)$.

2. STATEMENT OF MAIN RESULTS

Let $((X_i, Y_i))_{i \geq 1}$ be a stationary process of bivariate random vectors with marginal distribution function

$$F(x, y) = P(X_1 \leq x, Y_1 \leq y).$$

Throughout the article, we assume F to be Lipschitz continuous. For practical purposes this is fulfilled if F possesses a bounded density, but, for instance, $X = Y$ is also allowed. Kendall's rank correlation coefficient, also known as Kendall's tau, is a measure of dependence between the marginals X_1 and Y_1 . Kendall's tau is defined as

$$\tau = P((X' - X)(Y' - Y) > 0),$$

where (X, Y) and (X', Y') are two independent random variables with distribution function $F(x, y)$. The sample version of Kendall's tau is defined as

$$\hat{\tau}_n = \frac{1}{\binom{n}{2}} \#\{1 \leq i < j \leq n : (X_j - X_i)(Y_j - Y_i) > 0\}.$$

Remark 2.1.

- (i) The pairs (X_i, Y_i) and (X_j, Y_j) are called concordant, if $X_j - X_i$ has the same sign as $Y_j - Y_i$. Thus, $\hat{\tau}_n$ gives the fraction of concordant pairs among all pairs.
- (ii) Often, Kendall's tau is defined slightly differently as

$$\tilde{\tau} = P((X' - X)(Y' - Y) > 0) - P((X' - X)(Y' - Y) < 0),$$

which has the same range $[-1, 1]$ as Pearson's linear correlation. For continuous F , which we assume throughout, we have $\tilde{\tau} = 2\hat{\tau} - 1$, and the same one-to-one correspondence holds almost surely for the respective sample versions.

- (iii) The estimator $\hat{\tau}_n$ is a U -statistic with kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ defined as

$$(1) \quad h((x_1, y_1), (x_2, y_2)) = \mathbf{1}_{\{(x_2 - x_1)(y_2 - y_1) > 0\}}.$$

We will make use of this fact in our analysis of the asymptotic distribution of $\hat{\tau}_n$.

In this paper, we will study a test for change in the dependence structure of the marginals by the test statistic

$$\hat{T}_n = \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{\tau}_k - \hat{\tau}_n|.$$

Theorem 2.6 below gives the asymptotic distribution of the test statistic under the null hypothesis of no change in the correlation between the marginals of the pairs (X_i, Y_i) . Concerning the serial dependence structure of the process $((X_i, Y_i))_{i \geq 1}$, we assume that it is P -near epoch dependent (P -NED) on an absolutely regular process. The formal statement of this short range dependence assumption follows below. For simplicity of notation and consistency with large parts of the literature it is formulated for doubly infinite sequences of random vectors indexed by \mathbb{Z} . The observed data is then the positive branch of the doubly infinite sequence.

For two σ -fields \mathcal{A}, \mathcal{B} on the probability space (Ω, \mathcal{F}, P) , we define the absolute regularity coefficient

$$\beta(\mathcal{A}, \mathcal{B}) = E[\text{ess sup } \{|P(A|\mathcal{B}) - P(A)| : A \in \mathcal{A}\}].$$

The absolute regularity coefficient is a measure of dependence between the σ -fields \mathcal{A} and \mathcal{B} , it lies between 0 and 1, and equals 0 if \mathcal{A} and \mathcal{B} are independent.

Definition 2.2. Let $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ and $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ be q - and r -variate stochastic processes on (Ω, \mathcal{F}, P) , respectively, $q, r \geq 1$, such that the $(q + r)$ -variate process $((\mathbf{X}_n, \mathbf{Z}_n))_{n \in \mathbb{Z}}$ is stationary. For $k \leq n$, let $\mathcal{F}_k^n = \sigma(\mathbf{Z}_k, \dots, \mathbf{Z}_n)$, where also $k = -\infty$ and $n = \infty$ are permitted.

- (i) The process $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ is called *absolutely regular* if the absolute regularity coefficients

$$\beta_k = \beta(\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty), \quad k \geq 1,$$

converge to zero as $k \rightarrow \infty$.

- (ii) The process $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is called L_p near epoch dependent (L_p -NED), $p \geq 1$, on the process $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ if the approximating constants

$$a_{p,k} = \left(E \left| \mathbf{X}_0 - E(\mathbf{X}_0 | \mathcal{F}_{-k}^k) \right|_p^p \right)^{\frac{1}{p}}, \quad k \geq 1,$$

converge to zero as $k \rightarrow \infty$.

- (iii) The process $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is called *near epoch dependent in probability* or short *P-near epoch dependent* (*P-NED*) on the process $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ if there is a sequence of approximating constants $(a_k)_{k \in \mathbb{N}}$ with $a_k \rightarrow 0$ as $k \rightarrow \infty$, a sequence of functions $f_k : \mathbb{R}^{r \times (2k+1)} \rightarrow \mathbb{R}^q$, $k \in \mathbb{N}$, and a non-increasing function $\Phi : (0, \infty) \rightarrow (0, \infty)$ such that

$$(2) \quad P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon) \leq a_k \Phi(\varepsilon)$$

for all $k \in \mathbb{N}$ and $\varepsilon > 0$.

Remark 2.3.

- (i) The usual L_p -near epoch dependence, $p \geq 1$, is of lesser interest in the following. It is mentioned primarily to be put in contrast to *P*-near epoch dependence. It also appears in the proofs, where we make use of results formulated for L_1 - and L_2 -NED sequences. The main connection between the different approximation concepts is given by Lemma 2.5 below.
- (ii) The *P*-NED condition is equivalent to convergence in probability of $f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)$ to \mathbf{X}_0 for $k \rightarrow \infty$. If the latter is true, i.e., if

$$\varepsilon_k = \inf \left\{ \varepsilon \mid P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon) \leq \varepsilon \right\} \rightarrow 0 \quad (k \rightarrow \infty),$$

then setting $a_k = \varepsilon_k$ and

$$\Phi(\varepsilon) = \left(\sup_{l \leq k} \frac{P(|\mathbf{X}_0 - f_l(\mathbf{Z}_{-l}, \dots, \mathbf{Z}_l)|_1 > \varepsilon_k)}{\varepsilon_l} \right) \vee 1 \quad \text{for } \varepsilon \in [\varepsilon_k, \varepsilon_{k-1})$$

fulfills (2). The requirement of the bound on $P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon)$ in (2) to factorize into an ε -part and a k -part is not a restriction, but merely facilitates rate computations.

- (iii) Similar conditions that embody the idea of approximating functionals in a probability sense are *S*-mixing considered by Berkes, Hörmann, and Schauer (2009) and the L_0 -approximability of Pötscher and Prucha (1997, Chapter 6), who refer to the stochastic stability concept of Bierens (1981).
- (iv) The terminology *near epoch dependence* may be interpreted as follows: Near epoch dependence (in either definition) implies that there exists a measurable function $f : \mathbb{R}^{d \times \mathbb{N}} \rightarrow \mathbb{R}^d$ such that $\mathbf{X}_0 = f((\mathbf{Z}_k)_{k \in \mathbb{Z}})$ and by the stationarity of $((\mathbf{X}_n, \mathbf{Z}_n))_{n \in \mathbb{Z}}$ also $\mathbf{X}_n = f((\mathbf{Z}_{n+k})_{k \in \mathbb{Z}})$ for all $n \in \mathbb{Z}$. Thus in principle, \mathbf{X}_n depends on the whole process $(\mathbf{Z}_{n+k})_{k \in \mathbb{Z}}$, but the NED condition ensures that the dependence vanishes for the distant past and future, and \mathbf{X}_n primarily depends on the “near epoch” of \mathbf{Z}_n . An alternative terminology is “ $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is an approximating functional of $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ ” (where, as for NED, it remains to specify in which sense the approximation is meant).

We choose to consider *P*-near epoch dependence instead of the more frequently considered L_p near epoch dependence, $p \leq 1$, since we particularly want to analyze heavy tailed data and do not want to assume the existence of even first moments. The *P*-NED condition is a weaker assumption than L_p -NED, see Lemma 2.5 below, and the main results of this paper, in particular Theorems 2.6, 2.9, 3.2 and 3.5, hold for L_p -NED sequences as well. On the other hand, the *P*-NED condition substantially enlarges the class of processes for which the condition is easily checked by many heavy-tailed distributions. An example is given in the following lemma.

Lemma 2.4. *Let $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ be an i.i.d. sequence of \mathbb{R}^q -valued with $P(|\mathbf{Z}_n|_1 \geq t) \leq Ct^{-\alpha}$ for some $\alpha, C > 0$. Then for $a \in (-1, 1)$, the autoregressive process*

$$\mathbf{X}_n = \sum_{k=0}^{\infty} a^k \mathbf{Z}_{n-k}$$

is P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with $a_k = |a|^{\alpha k}$ and $\Phi(\varepsilon) = K\varepsilon^{-\alpha}$ for some constant $K > 0$.

Lemma 2.4 poses very weak conditions on the innovation distribution, in particular, the existence of a density is not required. This should be compared to analogous conditions for an AR(1) process to be mixing, see e.g. Withers (1981). All standard examples of discrete or heavy-tailed (e.g. Pareto, Cauchy, geometric) distributions are included here. The next lemma connects P -NED and L_p -NED.

Lemma 2.5. *Let $((\mathbf{X}_n, \mathbf{Z}_n))_{n \in \mathbb{Z}}$ be as in Definition 2.2.*

- (i) *If $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with functions Φ and f_k , $k \in \mathbb{N}$, and approximating constants $(a_k)_{k \in \mathbb{N}}$, and $g : \mathbb{R}^r \rightarrow \mathbb{R}^q$ is a Lipschitz continuous function with Lipschitz constant L , then the process $(g(\mathbf{X}_n))_{n \in \mathbb{Z}}$ is P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with functions $\tilde{\Phi}(\varepsilon) = \Phi(\varepsilon/L)$ and $g \circ f_k$, $k \in \mathbb{N}$, and the same approximating constants $(a_k)_{k \in \mathbb{N}}$.*
- (ii) *Let $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ be bounded and P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with functions Φ and f_k , $k \in \mathbb{N}$, and approximating constants $(a_k)_{k \in \mathbb{N}}$. Then $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is L_p -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ for any $p \geq 1$. If there is further a sequence $(s_k)_{k \in \mathbb{N}}$ of non-negative numbers such that*

$$(3) \quad a_k \Phi(s_k) = O(s_k) \quad (k \rightarrow \infty),$$

then the L_p -NED approximating constants $(a_{p,k})_{k \in \mathbb{N}}$ of $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ satisfy $a_{p,k}^p = O(s_k)$ for $k \rightarrow \infty$.

- (iii) *Let $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ be L_p -NED, $p \geq 1$, on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with approximating constants $(a_{p,k})_{k \in \mathbb{N}}$. Then $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$. If there is further a non-increasing function $\Phi : (0, \infty) \rightarrow (0, \infty)$ and a sequence $(s_k)_{k \in \mathbb{N}}$ of non-negative numbers converging to zero that satisfy*

$$\Phi(\varepsilon) s_k \geq \left(\frac{a_{p,k}}{\varepsilon} \right)^p,$$

then $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$ with approximation constants $(s_k)_{k \in \mathbb{N}}$ and function Φ . The functions f_k can be chosen as $f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) = E(\mathbf{X}_0 | \mathcal{F}_{-k}^k)$, $k \in \mathbb{N}$.

Since Φ is non-increasing, condition (3) puts an upper bound on the speed of decay of $(s_k)_{k \in \mathbb{N}}$ in the sense that, if (3) is fulfilled by some sequence $(s_k)_{k \in \mathbb{N}}$, then it is also fulfilled by any sequence $(\tilde{s}_k)_{k \in \mathbb{N}}$ for which $\tilde{s}_k \leq s_k$ for all k larger than some $n \in \mathbb{N}$. We are now ready to formulate our first main theorem, which concerns the asymptotic behavior of the test statistic \hat{T}_n under the null hypothesis of stationarity.

Theorem 2.6. *Let $(X_i, Y_i)_{i \geq 1}$ be a two-dimensional, stationary process that is P -NED with approximating constants $(a_k)_{k \geq 1}$ and function Φ on an absolutely regular process with absolute regularity coefficients $(\beta_k)_{k \geq 1}$ such that*

$$(4) \quad a_k \Phi(k^{-(3+\varepsilon)}) = O(k^{-(3+\varepsilon)}) \quad \text{and} \quad \beta_k = O(k^{-(1+\varepsilon)})$$

for some $\varepsilon > 0$. Then

$$(5) \quad \hat{T}_n \xrightarrow{\mathcal{D}} 2D \sup_{0 \leq \lambda \leq 1} |B(\lambda)|,$$

where $(B(\lambda))_{0 \leq \lambda \leq 1}$ denotes a Brownian bridge,

$$(6) \quad D^2 = \text{Var}(\psi(X_1, Y_1)) + 2 \sum_{j=2}^{\infty} \text{Cov}(\psi(X_1, Y_1), \psi(X_j, Y_j))$$

and $\psi(x, y) = 2F(x, y) - F_X(x) - F_Y(y) + 1 - \tau$.

Remark 2.7. The distribution of $\sup_{0 \leq \lambda \leq 1} |B(\lambda)|$ is known and often referred to as Kolmogorov distribution. It is the limiting distribution of the two-sided Kolmogorov–Smirnov test statistic. Its cdf is

$$(7) \quad F_K(x) = 1 - \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}, \quad x > 0.$$

In order to carry out the test in practice we need an estimate of D^2 . In principle, any consistent estimate \hat{D}_n^2 of D^2 can be used to obtain the convergence result (10), and the choice of the estimator certainly depends on the data situation. We follow in our proposal the kernel based estimation technique by de Jong and Davidson (2000), who prove asymptotic results for L_2 -NED sequences. Let \mathbb{F}_n , $\mathbb{F}_{X,n}$ and $\mathbb{F}_{Y,n}$ be the empirical distribution functions of $((X_i, Y_i))_{i=1, \dots, n}$, $(X_i)_{i=1, \dots, n}$ and $(Y_i)_{i=1, \dots, n}$, respectively, and

$$\hat{\psi}_{n,i} = 2\mathbb{F}_n(X_i, Y_i) - \mathbb{F}_{X,n}(X_i) - \mathbb{F}_{Y,n}(Y_i) + 1 - \hat{\tau}_n.$$

Then define

$$(8) \quad \hat{D}_n^2 = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_{n,i}^2 + \frac{2}{n} \sum_{j=1}^{n-1} \kappa\left(\frac{j}{b_n}\right) \sum_{i=1}^{n-j} \hat{\psi}_{n,i} \hat{\psi}_{n,i+j},$$

where $\kappa : [0, \infty) \rightarrow [-1, 1]$ is a kernel function satisfying Assumption 2.8 below and b_n is a bandwidth parameter depending on n . Assumption 2.8 mainly coincides with Assumption 1 of de Jong and Davidson (2000).

Assumption 2.8. *The kernel function $\kappa : [0, \infty) \rightarrow [-1, 1]$ with $\kappa(0) = 1$ is continuous at 0 and at all but a finite number of points. Furthermore, $|\kappa|$ is dominated by a non-increasing, integrable function and*

$$\int_{[0, \infty)} \left| \int_{[0, \infty)} \kappa(t) \cos(xt) dt \right| dx < \infty.$$

Theorem 2.9. *Let $(X_i, Y_i)_{i \geq 1}$ be a two-dimensional, stationary process that is P -NED with approximating constants $(a_k)_{k \geq 1}$ on an absolutely regular process and absolute regularity coefficients $(\beta_k)_{k \geq 1}$ such that*

$$(9) \quad a_k \Phi(k^{-(12+\varepsilon)}) = O(k^{-(12+\varepsilon)}) \quad \text{and} \quad \beta_k = O(k^{-(9+\varepsilon)})$$

for some $\varepsilon > 0$. Let furthermore κ be a kernel function satisfying Assumption 2.8 and $(b_n)_{n \in \mathbb{N}}$ be a non-decreasing sequence of natural numbers such that $b_n \rightarrow \infty$ and $b_n = o(n^{-1/2})$. Then

$$(10) \quad \frac{\hat{T}_n}{2\hat{D}_n} \xrightarrow{\mathcal{D}} \sup_{0 \leq \lambda \leq 1} |B(\lambda)|,$$

where $(B(\lambda))_{0 \leq \lambda \leq 1}$ is, as before, a Brownian bridge.

3. INVARIANCE PRINCIPLES FOR P -NED SEQUENCES

The proofs of the results of the previous section are based on two fundamental functional limit theorems for weakly dependent data. The key ingredient of the proof of Theorem 2.6 is Theorem 3.2 below, an invariance principle for U -statistics. Later in this section we state in Theorem 3.5 a multivariate empirical process invariance principle, which is essential for the proof of Theorem 2.9.

Let $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ denote a symmetric kernel, and let $(\mathbf{X}_i)_{i \geq 1}$ be a d -dimensional, stationary stochastic process with marginal distribution function $F : \mathbb{R}^d \rightarrow [0, 1]$. We define the U -statistic U_n as

$$U_n = U_n(g) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g(\mathbf{X}_i, \mathbf{X}_j).$$

In order to analyze the asymptotic distribution of U_n , we introduce the Hoeffding decomposition. Define

$$\begin{aligned} \theta &= Eg(\mathbf{X}, \mathbf{X}'), \\ (11) \quad g_1(\mathbf{x}_1) &= E(g(\mathbf{x}_1, \mathbf{X})) - \theta, \\ g_2(\mathbf{x}_1, \mathbf{x}_2) &= g(\mathbf{x}_1, \mathbf{x}_2) - g_1(\mathbf{x}_1) - g_1(\mathbf{x}_2) - \theta, \end{aligned}$$

where \mathbf{X} and \mathbf{X}' are independent random variables that each have the same distribution as \mathbf{X}_1 . Note that by definition, we get

$$g(\mathbf{x}_1, \mathbf{x}_2) = \theta + g_1(\mathbf{x}_1) + g_1(\mathbf{x}_2) + g_2(\mathbf{x}_1, \mathbf{x}_2).$$

The Hoeffding decomposition of U_n is given by

$$U_n(g) = \theta + \frac{2}{n} \sum_{i=1}^n g_1(\mathbf{X}_i) + \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} g_2(\mathbf{X}_i, \mathbf{X}_j).$$

For the U -statistic invariance principle to hold, we require the kernel g to fulfill the following regularity condition.

Definition 3.1. The kernel g satisfies the variation condition on $(\mathbf{X}_i)_{i \leq 1}$ if there exist constants $L, \varepsilon_0 > 0$ such that

$$E \left[\sup_{|(\mathbf{x}, \mathbf{x}') - (\mathbf{X}, \mathbf{X}')|_2 \leq \varepsilon} |g(\mathbf{x}, \mathbf{x}') - g(\mathbf{X}, \mathbf{X}')| \right] \leq L\varepsilon,$$

for all $\varepsilon \in (0, \varepsilon_0)$, where \mathbf{X} and \mathbf{X}' are independent random variables, identically distributed as \mathbf{X}_1 .

Theorem 3.2. Let $(\mathbf{X}_i)_{i \geq 1}$ be a d -dimensional, bounded, stationary process that is P -NED with approximating constants $(a_k)_{k \geq 1}$ on an absolutely regular process with coefficients $(\beta_k)_{k \geq 1}$ satisfying (4). Furthermore, let $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded kernel satisfying the variation condition. Then

$$(12) \quad (\sqrt{n} \lambda (U_{[n\lambda]} - \theta))_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} (2\sigma W(\lambda))_{0 \leq \lambda \leq 1},$$

where

$$(13) \quad \sigma^2 = \text{Var}(g_1(\mathbf{X}_1)) + 2 \sum_{i=2}^{\infty} \text{Cov}(g_1(\mathbf{X}_1), g_1(\mathbf{X}_i)).$$

and $(W(\lambda))_{0 \leq \lambda \leq 1}$ denotes a standard Brownian motion.

Remark 3.3.

- (i) Weak convergence in (12) is in the space $D([0, 1])$, equipped with the Skorokhod metric. Alternatively, one may consider a linearly interpolated version, i.e.

$$W_n(\lambda) = \begin{cases} \sqrt{n} \frac{k}{n} (U_k - \theta) & \text{for } \lambda = \frac{k}{n} \\ \text{linearly interpolated} & \text{in between.} \end{cases}$$

Then $(W_n(\lambda))_{0 \leq \lambda \leq 1}$ converges in distribution to $(2\sigma W(\lambda))_{0 \leq \lambda \leq 1}$, in the space $C([0, 1])$.

- (ii) The formulation of this invariance principle is specifically tailored to the needs of our application and does not strive for utmost generality. In particular the boundedness of g is a rather strict requirement that can be relaxed to a uniform bound on the $(2 + \eta)$ -moments of $g(\mathbf{X}_1, \mathbf{X}_k)$ for some $\eta > 0$. Contrary to the independent case, the existence of exactly second moments is generally not enough for series of dependent random variables. However, the relaxation of the boundedness must be paid for by a faster decay of the β_k and a_k coefficients. For details see Wendler (2011, Chap. 3).
- (iii) Denker and Keller (1986) proved a central limit theorem for U -statistics of approximating functionals of absolutely regular processes. A functional central limit theorem for U -statistics was established for absolutely regular processes by Yoshihara (1976) and by Denker and Keller (1983) under different sets of conditions. Theorem 3.2 extends these results to the much larger class of functionals of absolutely regular processes.

Corollary 3.4. *Under the assumptions of Theorem 3.2,*

$$(14) \quad (\sqrt{n} \lambda (U_{[n\lambda]} - U_n))_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} (2\sigma B(\lambda))_{0 \leq \lambda \leq 1},$$

where $(B(\lambda))_{0 \leq \lambda \leq 1}$ is a Brownian bridge and σ^2 defined in (13).

For the proof of Theorem 2.9 we require in particular that the distribution function F of (X_1, Y_1) is sufficiently well approximated by the empirical distribution function F_n (stemming from weakly dependent observations). This is ensured by the following functional limit theorem for the empirical process.

Theorem 3.5. *Let $(X_i, Y_i)_{i \geq 1}$ be a two-dimensional, stationary process that is P -NED with approximating constants $(a_k)_{k \geq 1}$ on an absolutely regular process and absolute regularity coefficients $(\beta_k)_{k \geq 1}$ satisfying (9). Then the empirical process*

$$(\sqrt{n}(F_n(s, t) - F(s, t)))_{s, t \in \mathbb{R}}$$

converges weakly to a centered Gaussian process $(W(s, t))_{s, t \in \mathbb{R}}$ with covariance function

$$E(W(s, t)W(s', t')) = \sum_{k=-\infty}^{\infty} \text{Cov}(\mathbf{1}_{\{X_0 \leq s, Y_0 \leq t\}}, \mathbf{1}_{\{X_k \leq s', Y_k \leq t'\}}).$$

4. CHANGE-POINT IDENTIFICATION

If the test rejects the null hypothesis of constant correlation, and if it is furthermore reasonable to assume that there is one sudden change-point, it is, of course, of interest to locate of this change-point. An intuitive estimator, which is common when dealing with CUSUM-type change-point tests, is the position at which the weighted correlation differences take their maximum, that is

$$\hat{b}_n = \arg \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{\tau}_k - \hat{\tau}_n|.$$

We will show in the following that this is indeed a reasonable estimator. We will assume that the following model holds.

Model 4.1 (Change-point model). Let $0 < b < 1$. For $n \in \mathbb{N}$ let $((X_{i,n}, Y_{i,n}))_{1 \leq i \leq [bn]}$ and $((X_{i,n}, Y_{i,n}))_{[bn]+1 \leq i \leq n}$ be two bivariate, stationary stochastic processes with marginal distribution functions F and G , respectively. Let furthermore $((X_{i,n}, Y_{i,n}))_{1 \leq i \leq n}$ be P -near epoch dependent¹ on an absolutely regular process with coefficients satisfying (9) uniformly for all n .

The goal is to estimate b . Let τ_F and τ_G denote Kendall's tau of F and G , respectively. Moreover, let $\tau_{FG} = Eh((X_1, Y_1), (X_2, Y_2))$, where $(X_1, Y_1) \sim F$ and $(X_2, Y_2) \sim G$ are independent.

Assumption 4.2. *The values $b, \tau_F, \tau_G, \tau_{FG} \in [0, 1]$ are such that $|c(\lambda)|$, $\lambda \in [0, 1]$, has a unique maximum at $\lambda = b$, where the function $c : [0, 1] \rightarrow \mathbb{R}$ is given by*

$$(15) \quad c(\lambda) = \begin{cases} [(1-b^2)\tau_F - (1-b)^2\tau_G - 2b(\lambda-b)\tau_{FG}] \lambda & \text{for } 0 \leq \lambda < b, \\ 2b(\tau_{FG} - \tau_G)(1-\lambda) + b^2(\tau_F + \tau_G - 2\tau_{FG}) \left(\frac{1}{\lambda} - \lambda\right) & \text{for } b \leq \lambda \leq 1. \end{cases}$$

Theorem 4.3. *If $((X_{i,n}, Y_{i,n}))_{1 \leq i \leq n, n \in \mathbb{N}}$ follows Model 4.1 and Assumption 4.2 holds, then*

$$\frac{\hat{b}_n}{n} \xrightarrow{p} b$$

as $n \rightarrow \infty$.

Remark 4.4. *Assumption 4.2 is satisfied in “most” situations. It is true if $\tau_F \neq \tau_G$ and*

$$\frac{(1-b)^2}{2((1-b)^2 + b)} \leq \frac{\tau_{FG} - \tau_F}{\tau_G - \tau_F} < 1.$$

It is violated if

$$0 < \frac{\tau_{FG} - \tau_F}{\tau_G - \tau_F} < \frac{(1-b)^2}{2((1-b)^2 + b)},$$

i.e., if τ_{FG} is too close to τ_F . It is an open research question which values of τ_{FG} are possible for given τ_F and τ_G , and in particular if τ_{FG} may at all lie outside the interval $[\tau_F, \tau_G]$.

5. PROPERTIES OF THE TEST AND COMPARISON TO PREVIOUS PROPOSALS

Wied, Krämer, and Dehling (2012) propose a test for constant correlation based on Pearson's linear correlation coefficient

$$\hat{\varrho}_k = \frac{\sum_{i=1}^k (X_i - \bar{X}_k)(Y_i - \bar{Y}_k)}{\left(\sum_{i=1}^k (X_i - \bar{X}_k)^2\right)^{1/2} \left(\sum_{i=1}^k (Y_i - \bar{Y}_k)^2\right)^{1/2}}.$$

The test statistic

$$\hat{T}_{\varrho,n} = \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{\varrho}_k - \hat{\varrho}_n|$$

is shown to converge in distribution to $D_\varrho \sup_{0 \leq t \leq 1} |B(t)|$ as $n \rightarrow \infty$ for NED sequences under the null hypothesis of no change in correlation, where $B(t)$ is a Brownian bridge and

¹ For non-stationary processes the short-range dependence conditions have to be formulated slightly more general than in Definition 2.2. The absolute regularity coefficients $(\beta_k)_{k \in \mathbb{N}}$ are defined as $\beta_k = \sup_{t \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+k}^\infty)$, and the P -NED approximation coefficients $(a_k)_{k \in \mathbb{N}}$ must satisfy

$$\sup_{t \in \mathbb{Z}} P(|\mathbf{X}_t - f_{k,t}(\mathbf{Z}_{t-k}, \dots, \mathbf{Z}_{t+k})|_1 > \varepsilon) \leq a_k \Phi_t(\varepsilon),$$

where the functions $f_{k,t}$ and Φ_t may also depend on t . The underlying process $(\mathbf{Z}_t)_{t \in \mathbb{Z}}$ is not required to be stationary.

D_ϱ an appropriate scaling factor. This test, devised for normal data, shall serve as the main benchmark method for our test. Our motivation for using Kendall's tau is the wish to efficiently detect structural changes in arbitrarily heavy-tailed and potentially contaminated data. Both tests are constructed in a similar way. The differences between the two tests are largely due to the different properties of the estimators $\hat{\varrho}_k$ and $\hat{\tau}_k$. It is therefore worthwhile to have a closer look at these two correlation measures.

The copula of the two-dimensional distribution function F is

$$C : [0, 1]^2 \rightarrow [0, 1] : (x, y) \mapsto F(F_X^{-1}(x), F_Y^{-1}(y)),$$

thus $F(x, y) = C(F_X(x), F_Y(y))$. Kendall's tau is a function of the copula,

$$(16) \quad \tau = 2 \int_{[0,1]^2} C(u, v) dC(u, v) = 2EC(U, V) = 2EF(X, Y),$$

where U, V are two uniformly on $[0, 1]$ distributed random variables with joint cdf C , e.g., $U = F_X(X)$ and $V = F_Y(Y)$, cf. Nelsen (2006, Chap. 5). The distribution function of $C(U, V)$ is also called the Kendall distribution function of the copula C . Kendall's tau consequently is invariant with respect to monotonic, componentwise transformations of F . Furthermore, $\hat{\tau}_n$ is asymptotically normal at the \sqrt{n} rate with asymptotic variance

$$(17) \quad ASV(\hat{\tau}_n) = 4\text{Var}(\psi(X, Y)) = 4\text{Var}(2C(U, V) - U - V)$$

for i.i.d. observations of any continuous distribution F , where $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ is defined in Theorem 2.6. Note that, instead of (16) and (17), the corresponding values for $\tilde{\tau} = 2\hat{\tau} - 1$ are usually given in the literature. Thus, no matter how large the tails of the distribution F are, as long as the marginals are joined by a Gauss copula, the asymptotic variance of $\hat{\tau}_n$, and hence the asymptotic distribution of $\hat{T}_{\tau,n}$, are the same as in the normal model.

A very popular class of multivariate distributions, which offers a convenient way of modeling data with arbitrarily heavy tails, is the elliptical model. A two-dimensional, continuous, centered, elliptical distribution F has a density f of the form

$$(18) \quad f(\mathbf{x}) = \det(S)^{-1/2} \gamma(\mathbf{x}^T S^{-1} \mathbf{x}), \quad \mathbf{x} = (x, y)^T \in \mathbb{R}^2,$$

where

$$S = \begin{pmatrix} s_{1,1} & s_{1,2} \\ s_{1,2} & s_{2,2} \end{pmatrix}$$

is a symmetric, positive definite matrix and $\gamma : [0, \infty) \rightarrow [0, \infty)$ a univariate function. We use the notation $\mathcal{E}_2(\mathbf{0}, S)$ for this class of distributions. If the second moments of F are finite, then

$$(19) \quad \varrho = \frac{s_{1,2}}{\sqrt{s_{1,1}s_{2,2}}}$$

is Pearson's linear correlation coefficient. Otherwise we use (19) as definition for the generalized linear correlation coefficient of the elliptical distribution F . In the elliptical model there is a one-to-one correspondence between ϱ and τ :

$$\tau = \frac{1}{\pi} \arcsin(\varrho) + \frac{1}{2}, \quad -1 \leq \varrho \leq 1.$$

Thus by letting

$$\hat{\varrho}_{\tau,n} = \sin(\pi(\hat{\tau}_n - 1/2)),$$

the estimators $\hat{\varrho}_n$ and $\hat{\varrho}_{\tau,n}$ are both Fisher-consistent for the same quantity ϱ , cf. (19), in the elliptical model. Comparing their asymptotic variances allows a prognosis concerning the efficiency relation of the corresponding change-point tests.

If 4th moments of $F \in \mathcal{E}_2(\mathbf{0}, S)$ are finite, then the asymptotic variance of $\hat{\varrho}_n$, computed from independent realizations of F , is

$$(20) \quad ASV(\hat{\varrho}_n) = (1 + \kappa/3)(1 - \varrho^2)^2,$$

where $\kappa = E(X^4)/(E(X^2))^2 - 3$ is the excess kurtosis of any component of $\mathbf{X} = (X, Y)$, where $\mathbf{X} \sim F$. At the two-dimensional normal distribution, i.e. for $\gamma(x) = \frac{1}{2\pi}e^{-x^2/2}$, the kurtosis κ is equal to zero. For the two-dimensional t -distribution with ν degrees of freedom (denoted by $t_{2,\nu}$ in the following), we have $\gamma(x) = c_\nu(1 - x/\nu)^{-\frac{\nu+2}{2}}$ for some normalizing constant c_ν and $\kappa = 6/(\nu - 4)$ for $\nu \geq 5$. The asymptotic variance of $\hat{\varrho}_{\tau,n}$ at the normal model is

$$(21) \quad ASV(\hat{\varrho}_{\tau,n}) = (1 - \varrho^2)\left(\frac{\pi^2}{9} - 4 \arcsin^2\left(\frac{\varrho}{2}\right)\right),$$

see Croux and Dehon (2010). For example, for two independent random variables we have $ASV(\hat{\varrho}_n) = 1$ and $ASV(\hat{\varrho}_{\tau,n}) = \pi^2/9 = 1.097$. For uncorrelated, jointly $t_{2,\nu}$ distributed random variables, Dengler (2010) gives the following expression for the asymptotic variance of $\hat{\varrho}_{\tau,n}$,

$$ASV(\hat{\varrho}_{\tau,n}) = \frac{32\Gamma(\frac{3\nu}{2})}{\pi^2\Gamma^3(\frac{\nu}{2})} \int_0^\infty u^{\nu-1} \arctan^2(u) \int_0^1 t^{\nu-1}(1-t)^{\nu-1}(u^2+t)^{-\nu} dt du,$$

and derives explicit expressions. The asymptotic variance $ASV(\hat{\varrho}_{\tau,n})$ is a decreasing function of ν , it equals 1.922 and 1.296 for $\nu = 1$ and $\nu = 5$, respectively, and is smaller than $ASV(\hat{\varrho}_n)$ for $\nu \leq 16$.

The maximum likelihood estimator $\hat{\varrho}_{t(\nu),n}$ of ϱ at the $t_{2,\nu}$ distribution has asymptotic variance

$$ASV(\hat{\varrho}_{t(\nu),n}) = \frac{\nu+4}{\nu+2}(1 - \varrho^2)^2, \quad \nu \geq 1,$$

which, for $\varrho = 0$, is equal to 1.667 for $\nu = 1$ and 1.286 for $\nu = 5$, see Bilodeau and Brenner (1999, p. 221). We note the remarkable fact that for all two-dimensional, uncorrelated t - and normal distributions the asymptotic relative efficiency of Kendall's tau with respect to the respective MLE is above 90% for $\nu \geq 2$. It is more than 99% at an uncorrelated $t_{2,5}$ distribution.

The other popular nonparametric correlation measure, Spearman's rho, which is often considered along with Kendall's tau, is defined as Pearson's linear correlation of the ranks of the data. It can be written as

$$\hat{r}_n = \frac{12}{(n-1)n(n+1)} \sum_{i=1}^n R_n(X_i)R_n(Y_i) - 3 \frac{n+1}{n-1},$$

where $R_n(X_i)$ denotes the rank of the i th observation X_i among X_1, \dots, X_n , likewise $R_n(Y_i)$. The population version of Spearman's rho,

$$(22) \quad s = 12 \int_{[0,1]^2} uv dC(u, v) - 3 = 12E(UV) - 3 = 12E(F_X(X)F_Y(Y)) - 3,$$

is also a function of the copula. Generally, Kendall's tau and Spearman's rho have similar statistical properties. See, e.g., Nelsen (2006, Chap. 5) for details on their relationship. Croux and Dehon (2010) compare both with respect to robustness and efficiency at the normal model and arrive at the conclusion, that in both respects their performance is comparable, but Kendall's tau is slightly favorable. Wied, Dehling, van Kampen, and Vogel (2011)

propose a non-parametric, robust change-point test for constant correlation for strongly mixing sequences that is closely related to Spearman's rho. They consider the test statistic

$$\hat{T}_{s,n} = \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{s}_k - \hat{s}_n|,$$

where

$$(23) \quad \hat{s}_k = \frac{12}{n^3} \sum_{i=1}^k R_n(X_i) R_n(Y_i) - 3 - \frac{12}{n}, \quad 1 \leq k \leq n.$$

The proof of its convergence is based on an invariance principle for the multivariate sequential empirical process by Rüschendorf (1976). Despite the mentioned practical parity of Kendall's tau and Spearman's rho, this test has a low efficiency compared to our Kendall's tau based test (see Section 6). The reason lies in the usage of $R_n(\cdot)$ instead of $R_k(\cdot)$ in (23). Spearman's rho is asymptotically equivalent to a U -statistic of order 3, see Moran (1948), and an asymptotic analysis of the related test statistic

$$\hat{T}_{r,n} = \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{r}_k - \hat{r}_n|$$

by means of U -statistics theory is mathematically much more involved. Since Spearman's rho, on the other hand, exhibits no pronounced advantage over Kendall's tau, we do not pursue this further here. Finally, we note that both estimators require a comparable computing effort. Both can be computed in $O(n \log n)$ time. Simple algorithms to compute the test statistics require $O(n^2)$.

6. SIMULATION RESULTS

In this section we give some numerical results, primarily addressing two questions: we want to (A) examine the goodness of the asymptotic approximation of the distribution of the test statistic for finite n under different dependence scenarios and (B) compare the performance of the test to the proposals by Wied et al. (2012, 2011) with respect to efficiency and robustness. Furthermore (C) we demonstrate the applicability of our test at an example which is neither absolutely regular nor an L_p approximating functional, $p \geq 1$, of an absolutely regular process, but which is easily be shown to be P -NED on an i.i.d. process.

For objectives (A) and (B) we consider two simple models that substantially differ with respect to the strength of the serial dependence. Both fit into the framework of the following general model.

General model. The random vectors $(\delta_i, \varepsilon_i)$, $i \in \mathbb{Z}$, are independent and identically distributed, each having a bivariate, centered elliptical distribution $\mathcal{E}_2(\mathbf{0}, S)$ with

$$S = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix},$$

where the shape parameter $|\varrho| \leq 1$ is equal to the usual moment correlation if the second moments of $(\delta_1, \varepsilon_1)$ are finite. The series $((X_i, Y_i))_{i \in \mathbb{Z}}$ then follows the AR(1) process

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \varphi \begin{pmatrix} X_{i-1} \\ Y_{i-1} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix}, \quad i \in \mathbb{Z},$$

with AR parameter $|\varphi| < 1$.

Model 6.1. $\varphi = 0$.

Model 6.2. $\varphi = 0.8$.

The independent observations of Model 6.1 constitute the “good case” scenario (the best realistically assumable case), while Model 6.2 implements a dependence scenario with strong positive autocorrelations, where we expect the test to be less efficient.

Throughout we estimate the cumulated covariance D^2 by the estimator \hat{D}_n^2 given by (8), where we choose κ to be the quartic kernel

$$\kappa(x) = (1 - x^2)^2 \mathbf{1}_{[-1,1]}(x)$$

and the bandwidth $b_n = \lfloor 2n^{1/3} \rfloor$.

Objective (A). If we let the distribution of $(\delta_1, \varepsilon_1)$ be bivariate normal with $\varrho = 0$, the margins of (X_i, Y_i) are independent, $i \in \mathbb{Z}$, and, in both Models 6.1 and 6.2, D^2 can be given an explicit form. We have

$$D^2 = \frac{1}{36} = 0.02\bar{7}$$

for Model 6.1 and

$$D^2 = \frac{1}{36} + \frac{2}{\pi^2} \sum_{j=1}^{\infty} \arcsin^2\left(\frac{0.8^j}{2}\right) = 0.12099$$

for Model 6.2. The values can be deduced from Lemma D.1 in the appendix. Recall that $\arcsin(x)$ is very close to x for x close to zero. This allows us to compare how well $\hat{T}_n/(2\hat{D}_n)$ and $\hat{T}_n/(2D)$ are approximated by their common limit distribution. In Figures 1 and 2, the cdf F_K of the limiting Kolmogorov distribution, cf. (7), is plotted along with several empirical distribution functions, each based on 5000 repetitions. The left panels show empirical distribution functions of $\hat{T}_n/(2D)$ for different n and the right panels empirical cdf's of $\hat{T}_n/(2\hat{D}_n)$. In Figure 1, the results for Model 6.1 with $\varrho = 0$ are displayed. While for $n = 100$, $\hat{T}_n/(2D)$ is already very close to its limit distribution, there is some considerable bias for $\hat{T}_n/(2\hat{D}_n)$, which vanishes for $n \geq 500$. The results for Model 6.2 (also with $\varrho = 0$) in Figure 2 look different. The convergence of $\hat{T}_n/(2D)$ is much slower. For $n = 1000$ there is still a considerable gap between the empirical cdf and its limit. Somewhat surprising but good news is that $\hat{T}_n/(2\hat{D}_n)$ converges faster in this situation. Here the asymptotics are usable for $n \geq 500$. For both models we ran simulations for $n = 10, 20, 50, 100, 500$ and 1000. For clarity of visual representation, some of the curves are omitted. Where the curves for $n = 500$ and $n = 1000$ are not displayed, they practically agree with the graph of F_K .

Objective (B). Guided by the observations above we use a fixed sample size of $n = 500$ for the efficiency comparison of our test to the previous proposals by Wied et al. (2012) and Wied et al. (2011). These tests are based on $\hat{T}_{\varrho,n}$ and $\hat{T}_{s,n}$, respectively (cf. Section 5) and referred to as *Pearson test* and *copula test* in the following. The variance estimation for these test statistics is done according to the authors' proposals, which both also implement kernel estimators following de Jong and Davidson (2000). We also take the quartic kernel and the bandwidth $b_n = \lfloor 2n^{1/3} \rfloor$.

For the first half of the data, we sample independent realizations $(\delta_i, \varepsilon_i)$, $i = 1, \dots, 250$, with correlation parameter $\varrho_1 = 0.4$. For the second half of the data, we use the correlation parameter ϱ_2 , for which we allow the values 0.4 (null hypothesis), 0.6, 0.8, 0.2, 0, -0.2 , -0.4 . Thus in Model 6.1, where $(X_i, Y_i) = (\delta_i, \varepsilon_i)$, we have a constant correlation of 0.4 at the beginning and then a sudden jump. In Model 6.2, there is an abrupt jump in the correlation of the innovations $(\delta_i, \varepsilon_i)$ at time $n/2$, which means a gradual but quick change in the correlation of the observed process (X_i, Y_i) . Note that the stationary processes $((\delta_i, \varepsilon_i))_{i \in \mathbb{Z}}$ and $((X_i, Y_i))_{i \in \mathbb{Z}}$ have the same correlation.

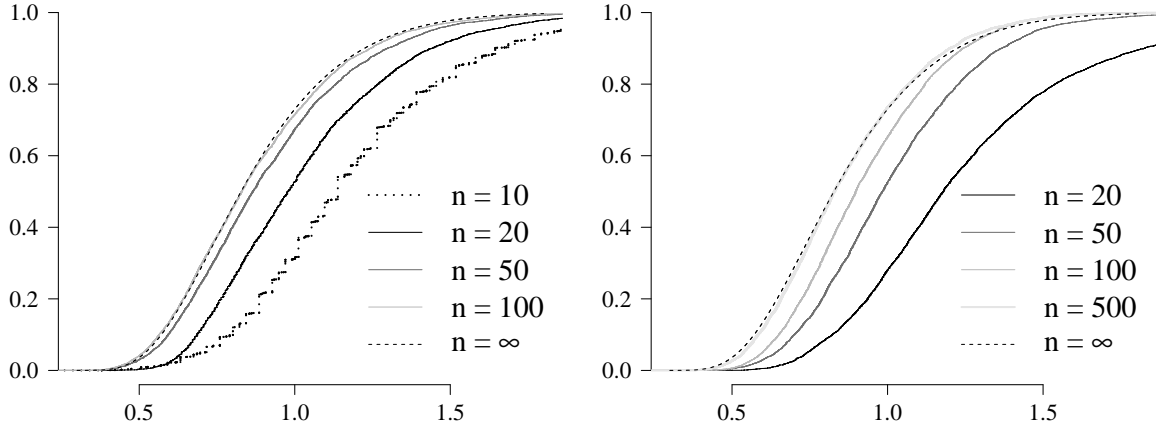


FIGURE 1. Convergence of $\hat{T}_n/(2D)$ (left) and $\hat{T}_n/(2D_n)$ (right) to the Kolmogorov distribution for serial independence; empirical cdfs based on 5000 repetitions.

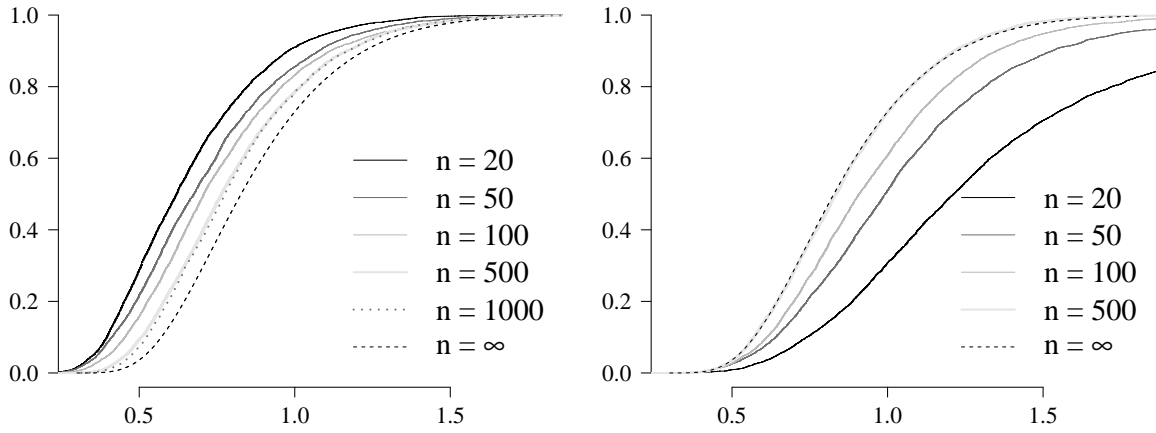


FIGURE 2. Convergence of $\hat{T}_n/(2D)$ (left) and $\hat{T}_n/(2D_n)$ (right) to the Kolmogorov distribution for independent Gaussian AR(1) processes with AR parameter $\varphi = 0.8$; empirical cdfs based on 5000 repetitions.

We consider five different elliptical distributions for $(\delta_i, \varepsilon_i)$: the bivariate normal distribution and bivariate t -distribution with 20, 5, 3 and 1 degrees of freedom. The t_{20} distributions has slightly heavier tails than the normal, whereas t_5 , t_3 and t_1 serve as examples of very heavy tailed distributions. The t_ν distribution possesses finite moments of order $\nu - 1$, but no higher whole number. Thus t_5 is the “smallest t -distribution” for which the Pearson test by Wied et al. (2012) works, and t_3 is the “smallest t -distribution” for which Pearson’s moment correlation is defined.

For each combination of model, jump height and marginal distribution we generate 1000 samples and compute the three test statistics from each sample. The observed rejection frequencies at the significance level .05 are given in Tables 1 and 2 for Models 6.1 and 6.2, respectively. At Table 1 we note the following.

- (1) The Pearson test is slightly better than the Kendall test at the normal distribution. Both tests lose power with increasing tails, but the loss is much smaller for the Kendall test.

TABLE 1. Efficiency comparison of three correlation change-point tests under Model 6.1. Different marginal distributions, 500 observations, different jump sizes in the middle of the sample. Empirical rejection frequencies at the asymptotic .05 level based on 1000 repetitions.

Jump at $n/2$:		none	-.2	+.2	-.4	+.4	-.6	-.8
Distribution	Test							
normal	Pearson	.04	.46	.70	.97	1.00	1.00	1.00
	copula	.04	.06	.07	.22	.20	.47	.78
	Kendall	.05	.44	.65	.96	1.00	1.00	1.00
t_{20}	Pearson	.04	.42	.65	.97	1.00	1.00	1.00
	copula	.03	.07	.08	.22	.20	.47	.77
	Kendall	.04	.46	.63	.97	1.00	1.00	1.00
t_5	Pearson	.04	.24	.41	.73	.95	.95	.98
	copula	.04	.08	.08	.22	.20	.46	.76
	Kendall	.04	.41	.55	.95	1.00	1.00	1.00
t_3	Pearson	.06	.14	.25	.39	.69	.64	.79
	copula	.03	.08	.08	.21	.18	.43	.72
	Kendall	.03	.39	.52	.91	1.00	1.00	1.00
t_1	Pearson	.47	.48	.50	.49	.56	.52	.51
	copula	.03	.06	.07	.17	.17	.38	.63
	Kendall	.04	.29	.38	.83	.98	1.00	1.00

For the t_{20} distribution, the results are comparable. The Kendall test is clearly better for heavier tails. These observations are fully in line with our expectations considering the efficiency comparison of the respective correlation measures in Section 5.

- (2) Throughout, the copula test has a very low power.
- (3) A positive jump (from the positive correlation $\varrho = 0.4$) is better detected by the Kendall and the Pearson test than a negative jump of the same height. This also to be expected considering that correlation measures generally have a smaller variance if the true absolute correlation is large, cp. also (20) and (21).
- (4) The Pearson test has no mathematical justification if fourth moments do not exist. For the t_3 distribution it gives nevertheless approximate results, whereas for the t_1 distribution it is completely useless.

Analyzing Table 2 we find that

- (5) the power of all tests is lower for the AR(1) process than in the independent case, and
- (6) the observations made at Table 1 concerning the comparison of the tests generally also apply here.

There are some differences, though.

- (7) For normal, t_{20} , t_5 and t_3 innovations, the performance of the Kendall and the Pearson test are rather similar. The effect of the heavy tails is less pronounced than in the independent case. This is not entirely surprising. In Model 6.2,

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \sum_{k=0}^{\infty} \varphi^k \begin{pmatrix} \delta_{i-k} \\ \varepsilon_{i-k} \end{pmatrix}$$

TABLE 2. Efficiency comparison of three correlation change-point tests. $((X_i, Y_i))_{i=1, \dots, n}$ AR(1) process with AR-parameter $\varphi = 0.8$. Different innovation distributions, 500 observations, several alternatives. Empirical rejection frequencies at the asymptotic .05 level based on 1000 repetitions.

Jump at $n/2$:		none	-.2	+.2	-.4	+.4	-.6	-.8
Distribution	Test							
normal	Pearson	.07	.13	.27	.45	.77	.78	.94
	copula	.05	.06	.04	.07	.07	.11	.18
	Kendall	.05	.14	.19	.46	.73	.79	.96
t_{20}	Pearson	.06	.11	.31	.41	.77	.77	.94
	copula	.05	.06	.06	.08	.08	.12	.20
	Kendall	.04	.13	.23	.42	.74	.79	.95
t_5	Pearson	.08	.12	.26	.34	.67	.67	.89
	copula	.06	.05	.06	.08	.09	.12	.18
	Kendall	.05	.15	.19	.40	.67	.73	.95
t_3	Pearson	.10	.11	.26	.25	.56	.50	.67
	copula	.06	.08	.06	.08	.09	.09	.17
	Kendall	.05	.12	.18	.34	.62	.67	.90
t_1	Pearson	.46	.49	.52	.50	.56	.50	.54
	copula	.08	.07	.08	.09	.11	.12	.14
	Kendall	.06	.12	.11	.18	.34	.34	.53

is a sum of independent random variables. Although the sum is generally not normal (the summands do not satisfy the Lindeberg condition) and does not possess any higher moments than $(\delta_i, \varepsilon_i)$ itself, it is, purely heuristically speaking, closer to a normal distribution than the innovations (if these possess finite second moments). Thus we expect the performance of the change-points tests to be closer to that at the normal model than it is the case in Model 6.1.

- (8) We have noted that positive jumps are generally better detected (starting from correlation 0.4). Furthermore, we note at both tables, but more clearly at Table 2, that the difference in power (positive jump vs. negative jump of equal height) is larger for the Pearson test than for the Kendall test. In Table 2 we even find that the Kendall test is always better at detecting negative jumps, where as Pearson is better at detecting the majority of the positive jumps. This is favorable for the Kendall test, since in practical situations one is much more likely to encounter a change in correlation from, say, 0 to 0.4 than from 0.4 to 0.8. This behavior is to be expected comparing (20) and (21): at the normal model the asymptotic relative efficiency of Kendall's tau with respect to Pearson's correlation coefficient increases as the absolute value of the true correlation decreases and reaches its maximum of $9/\pi^2$ at $\varrho = 0$. However, we have no apparent reason why the effect is much more pronounced in the AR(1) case than in the independent case.

The apparent advantages of the Pearson test for large non-zero correlations observed at Tables 1 and 2 must be put into perspective with the following: First, in Table 2 the Pearson test never keeps the 0.05 significance level under the null hypothesis. Adjusting the critical value accordingly will result in much lower rejection frequencies under the alternatives.

TABLE 3. Robustness comparison of correlation change-point tests. Serial independence, bivariate Gaussian distribution with $\rho = 0.4$. Different sizes and fractions of outliers; 500 observations. Rejection frequencies at the asymptotic .05 level based on 1000 repetitions.

Outlier magnitude:		$U(0, 5)$				$U(10, 100)$			
Outlier fraction (%):		1	2	5	10	1	2	5	10
Test:	Pearson	.09	.40	.99	1.00	1.00	1.00	1.00	1.00
	copula	.03	.04	.10	.31	.03	.06	.15	.52
	Kendall	.06	.11	.67	.99	.10	.25	.89	1.00

Second, ellipticity implies that any monotone dependence is basically linear. This intrinsic linearity of the elliptical model favors measures of linear dependence, but is a questionable assumption in practice. In particular, strong linear dependence is rarely encountered. Often, *monotone* dependence is interest rather than *linear* dependence. The prevalent use of linear correlation coefficients to measure monotone dependence is presumably due to their simplicity and amplified by their historical dominance. In alternative models, that may exhibit strong monotone but not necessarily linear dependence, the picture is much better for the Kendall test. We present results for elliptical distributions due to their widespread use.

We have noticed that in all simulations the copula test has a very low power and is outperformed by the Kendall test. Besides the applicability for heavy tailed data the second motivation for introducing both tests, copula and Kendall, is the lack of robustness of the Pearson test. The question remains how both test compare with respect to their robustness properties. We want to obtain a rough idea by simulating an outlier scenario. We use, as before, $n = 500$ and sample from the null hypothesis of Model 6.1 with Gaussian margins, i.e., (X_i, Y_i) are independent and have constant correlation 0.4. In the second half of the sample we randomly replace some observations by outliers. The outliers are of the form $(\xi_i, -\xi_i)$, where ξ_i is drawn from either $U(-5, 5)$ (small outliers) or from $U((-100, -10) \cup (10, 100))$ (large outliers), suggesting a strong negative correlation. The position of the outliers is a uniform sample without replacement from 251, ..., 500. Although size as well as position of the outliers are random, structure and placement are very unfavorable for the null hypothesis. This outlier scenario can be considered as a worst case, atypical for contaminated data encountered in practice. The results are given in Table 3. The Kendall test can cope with a few bad outliers of the described type, but gets considerably biased as the number of outliers increases. The copula test can cope very well even with a large fraction of severe outliers, but it is debatable whether this makes up for the low efficiencies reported in Tables 1 and 2.

Objective (C). The example processes considered so far are all absolutely regular with exponential decay of the mixing coefficients. However, Theorem 2.9 poses the much weaker condition on the data process to be P -near epoch dependent on an absolutely regular process. In the remainder of this section we study an example of a process which is neither absolutely regular itself nor L_1 -near epoch dependent on any process (in the sense of Definition 2.2 (ii)), but which can be treated by our general P -NED formulation.

Model 6.3. Let, as before, $((X_i, Y_i))_{i \in \mathbb{Z}}$ follow an AR(1) process

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \varphi \begin{pmatrix} X_{i-1} \\ Y_{i-1} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix}, \quad i \in \mathbb{Z},$$

TABLE 4. Heavy-tailed, non-mixing process (Model 6.3). 500 observations, several alternatives. Empirical rejection frequencies at the asymptotic .05 level based on 1000 repetitions.

Jump at $n/2$:	none	-.2	+.2	-.4	+.4	-.6	-.8
copula test:	.04	.03	.05	.09	.09	.19	.36
Kendall test:	.05	.38	.36	.89	.99	1.00	1.00

where the $(\delta_i, \varepsilon_i)$, $i \in \mathbb{Z}$, are independent and identically distributed. Now let $\varphi = 1/2$ and $(\delta_i, \varepsilon_i)$ have the following discrete distribution

$$P((\delta_i, \varepsilon_i) = (0, \tfrac{1}{2})) = P((\delta_i, \varepsilon_i) = (\tfrac{1}{2}, 0)) = (1 - \varrho)/4,$$

$$P((\delta_i, \varepsilon_i) = (0, 0)) = P((\delta_i, \varepsilon_i) = (\tfrac{1}{2}, \tfrac{1}{2})) = (1 + \varrho)/4.$$

The parameter $\varrho \in [-1, 1]$ is the moment correlation of this distribution. The Kendall's tau coefficient (the one symmetric around 0, cf. Remark 2.1 (ii)) of $(\delta_i, \varepsilon_i)$ is $\tilde{\tau} = \varrho/2$. The process $((X_i, Y_i))_{i \in \mathbb{Z}}$ is not strongly mixing (see e.g. Ibragimov and Linnik (1971), p. 360), and hence not absolutely regular, but by Lemma 2.4 it is P -NED on $((\delta_i, \varepsilon_i))_{i \in \mathbb{Z}}$ with exponentially decreasing approximation coefficients. The pair (X_i, Y_i) has the same moment correlation ϱ as $(\delta_i, \varepsilon_i)$, its Kendall rank correlation is $\tilde{\tau} = 2\varrho/(3 - \varrho^2)$, and the margins X_i and Y_i are uniformly distributed on $(0, 1)$. We simulate data from the process $((\tilde{X}_i, \tilde{Y}_i))_{i \in \mathbb{Z}}$ with

$$\tilde{X}_i = H(X_i), \quad \tilde{Y}_i = H(Y_i),$$

where H denotes the quantile function of a Pareto (type I) distribution with shape parameter $1/2$ and location parameter 1, i.e.

$$H(x) = \frac{1}{(1-x)^2}, \quad x \in [0, 1).$$

This strictly increasing transformation leaves the P -NED coefficients as well as Kendall's tau unchanged. The margins \tilde{X}_i, \tilde{Y}_i are Pareto distributed and have finite moments only up to order less than $1/2$.

The simulation set-up (including bandwidth and kernel for the variance estimation) is exactly the same as in Table 2. The sample size is always 500, the correlation parameter ϱ is equal to 0.4 for the first 250 innovations and then jumps by one of the values given in Table 4. The reported rejection frequencies of the tests are based on 1000 repetitions. The results are comparable to those in the heavy-tailed, elliptical i.i.d. case, cf. Table 1.

7. DATA EXAMPLES

Wied et al. (2012) analyze the dependence between the German stock index (DAX) and the Standard and Poor 500 (S&P 500). We apply the Kendall test and the Pearson test (with the same parameter choices for the variance estimation as in the simulation section) to the daily log returns of the two financial indices in the years 2006 through 2009 (1043 observations). The second half of this period covers what has been termed the Global Financial Crises. The processes $(\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|)_{k=1, \dots, n}$ and $(\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|)_{k=1, \dots, n}$ are depicted in Figure 3. Their maxima are the values of the test statistics of the Pearson and the Kendall test, respectively. Both tests give a p-value below 0.005, and both attain their maximum on July 14, 2008, at the height of the financial crisis. (Lehman Brothers filed for bankruptcy on September 14, 2008.) The tests behave similarly and their outcome supports the assumption that the

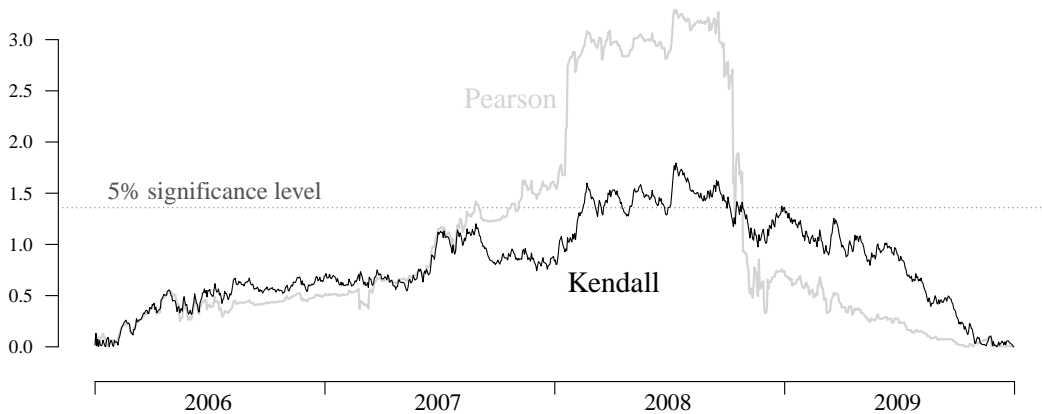


FIGURE 3. Processes $\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|$ (grey) and $\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|$ (black), $k = 1, \dots, n$, computed from log returns of DAX and S&P 500 between Jan 1, 2006, and Dec 31, 2009.

dependence between both indices considered can not be assumed to be identical before and during the financial crisis of 2008.

At a second data example, the difference between both tests become apparent. We consider the Dow Jones Industrial Average and the Nasdaq Composite in the years 1987 and 1988 (Figure 4). The most notable feature of both time series is the heavy loss on October 19, 1987, commonly known as Black Monday. Here we may ask in particular the question if the market conditions substantially changed after this date. Does the Black Monday constitute a break in the correlation between the two time series? The Pearson test reports a p-value indistinguishable from zero by machine accuracy. The underlying processes of the Pearson and the Kendall test are shown in Figure 5. The outcome of the Pearson test is determined by the peak on October 19, 1987, which is explained as follows. On October 19, both indices suffered heavy losses, suggesting a strong positive correlation of their log returns. The following day the Dow Jones recovered to some small degree, whereas the Nasdaq experienced an even larger drop, suggesting strong negative correlation. Thus the process of successive sample correlations jumps up and immediately down again.

The Kendall test gives a p-value of 0.24, indicating one can assume the correlation between the two time series to be constant over the observed time period. The empirical Kendall's tau is 0.52 prior to Black Monday, and 0.56 afterwards. Indeed, the market conditions turned out to be not much different from before, the DJIA even closed positive for 1987.

The strong impact of a few or even a single extreme observation on the Pearson test demonstrates once more the inappropriateness of the moment correlation for heavy-tailed data.

8. CONCLUSION

We have presented a fluctuation test for detecting changes in the dependence between two time series based on Kendall's rank correlation coefficient. We have demonstrated the non-inferiority of the test in terms of efficiency and the clear superiority in terms of robustness and applicability to a similar, previously proposed test, which is based on Pearson's moment correlation. To allow arbitrarily heavy-tailed data and very weak assumptions concerning the serial dependence, we have introduced the concept of near epoch dependence in probability.

We have studied the asymptotic behavior of the test statistic under stationarity by means of limit theorems for U -statistics for weakly dependent, stationary processes. Simulations

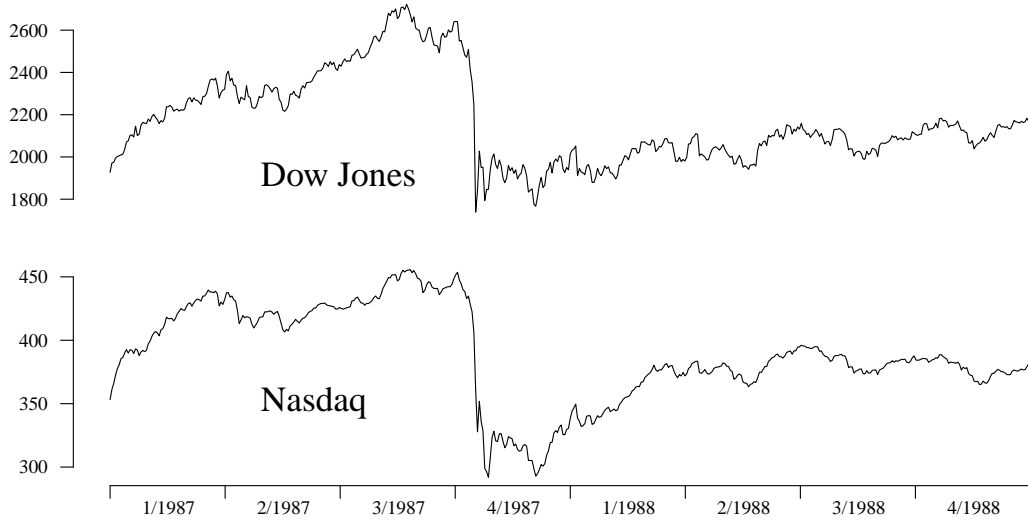


FIGURE 4. Daily closings of Dow Jones Industrial Average and Nasdaq Composite from Jan 1, 1987, to Dec 31, 1988.



FIGURE 5. Processes $\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|$ (grey) and $\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|$ (black), $k = 1, \dots, n$, computed from log returns of DJIA and Nasdaq Composite between Jan 1, 1987, and Dec 31, 1988.

show that the proposed test possesses a variety of advantageous features that have not been discussed in this paper. It has power against gradual or fluctuating changes in the correlation, not only sudden jumps, as presented in Section 6. It also exhibits a much better robustness against heteroscedasticity than the Pearson test.

However, a thorough, theoretical assessment of these properties as well as constructing tests that explicitly allow heteroscedasticity require the study of U -statistics at non-stationary sequences. This is mathematically rather challenging and – to our knowledge – not treated in the literature. It is certainly a topic of future research and goes beyond the scope of this paper.

A yet unsatisfactory property of the test is the lack of finite sample accuracy in the case of strong serial dependence. This can be overcome by bootstrapping the test statistic using a block bootstrap, but the theoretical justification for such a procedure, again, holds some considerable mathematical challenge.

Further future research directions include, e.g., the extension to more than two dimensions or guidelines for an on-line application of the test with results about the detection time of a

change. Another interesting research question, which is related to the one studied here, is to devise a robust test for detecting changes in the *coherence* of two time series. For example, a series of i.i.d. variables, shifted by one observation, is highly coherent to the original series, but our test, which only compares observations at the same time point, does not detect that type of dependence.

ACKNOWLEDGMENT

The authors wish to thank their colleague Roland Fried for several very stimulating discussions that motivated this paper. The research was supported in part by the Collaborative Research Grant 823 *Statistical modelling of nonlinear dynamic processes* of the German Research Foundation.

APPENDIX A. PROOFS OF SECTION 2

Proof of Lemma 2.4. Taking $f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) = \sum_{l=0}^k a^l \mathbf{Z}_{-l}$, we have to show that

$$P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon) = P\left(\left|\sum_{l=0}^{\infty} a^l \mathbf{Z}_l\right|_1 > \frac{\varepsilon}{|a|^{k+1}}\right) \leq K\varepsilon^{-\alpha}|a|^{\alpha k}.$$

Let $|a| < b < 1$. If, for all $l \in \mathbb{N}$,

$$|a^l \mathbf{Z}_l|_1 \leq (1-b)b^l \frac{\varepsilon}{|a|^{k+1}},$$

then

$$\left|\sum_{l=0}^{\infty} a^l \mathbf{Z}_l\right|_1 \leq \sum_{l=0}^{\infty} |a^l \mathbf{Z}_l|_1 \leq \frac{\varepsilon}{|a|^{k+1}}.$$

Hence

$$\begin{aligned} P\left(\left|\sum_{l=0}^{\infty} a^l \mathbf{Z}_l\right|_1 > \frac{\varepsilon}{|a|^{k+1}}\right) &\leq P\left(\bigcup_{l \in \mathbb{N}} \left\{|a^l \mathbf{Z}_l| > \frac{(1-b)b^l \varepsilon}{|a|^{k+1}}\right\}\right) \\ &\leq \sum_{l=0}^{\infty} P\left(|a^l \mathbf{Z}_l| > \frac{(1-b)b^l \varepsilon}{|a|^{k+1}}\right) \leq \sum_{l=0}^{\infty} C \left(\frac{(1-b)b^l \varepsilon}{|a|^{k+1+l}}\right)^{-\alpha} \\ &= C \left(\frac{\varepsilon(1-b)}{|a|^{k+1}}\right)^{-\alpha} \sum_{l=0}^{\infty} \left(\frac{|a|}{b}\right)^{\alpha l} \leq K\varepsilon^{-\alpha}|a|^{\alpha k} \end{aligned}$$

for some $K > 0$. □

Proof of Lemma 2.5. Part (i) is straightforward.

Part (ii): There are positive constants C_1, C_2 such that

$$(24) \quad a_{p,k}^p = E|\mathbf{X}_0 - E(\mathbf{X}_0 | \mathcal{F}_{-k}^k)|_p^p \leq C_1 E|\mathbf{X}_0 - E(\mathbf{X}_0 | \mathcal{F}_{-k}^k)|_1$$

$$(25) \quad \leq C_2 E|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1.$$

The first inequality (24) is due to the boundedness of \mathbf{X}_0 , and the constant C_1 depends on its bound and p . The second inequality (25) does generally not hold with $C_1 = C_2$. The conditional expectation provides the best L_2 approximation, but here we consider the L_1 distance. We may, however, argue as follows: By applying Jensen's inequality for the conditional expectation to the convex function $|\cdot|_1$ we obtain

$$|E(\mathbf{X}_0 | \mathcal{F}_{-k}^k) - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 \leq E\left(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 \middle| \mathcal{F}_{-k}^k\right),$$

and by taking the expectation of both sides we get

$$\begin{aligned} & E \left| \mathbf{X}_0 - E(\mathbf{X}_0 | \mathcal{F}_{-k}^k) \right|_1 \\ & \leq E \left| \mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) \right|_1 + E \left| f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) - E(\mathbf{X}_0 | \mathcal{F}_{-k}^k) \right|_1 \\ & \leq 2E \left| \mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) \right|_1. \end{aligned}$$

Now for any $\varepsilon > 0$ we have

$$\begin{aligned} & E \left| \mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) \right|_1 \\ & = E \left(\left| \mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) \right|_1 \mathbf{1}_{\{|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon\}} \right) \\ & \quad + E \left(\left| \mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) \right|_1 \mathbf{1}_{\{|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 \leq \varepsilon\}} \right) \\ & \leq C_3 P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon) + \varepsilon \leq C_3 \Phi(\varepsilon) a_k + \varepsilon \end{aligned}$$

Combining this with (25) we arrive at

$$a_{p,k}^p \leq C_2 C_3 \Phi(\varepsilon) a_k + C_2 \varepsilon$$

By first choosing ε sufficiently small and then k sufficiently large we can make the left-hand side arbitrarily small, and hence $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is L_p -NED, $p \geq 1$, on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$.

In particular, if condition (3) holds, we get by taking $\varepsilon = s_k$:

$$a_{p,k}^p \leq C_2 C_3 \Phi(s_k) a_k + C_2 s_k = O(s_k).$$

Part (iii): Letting $f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k) = E(\mathbf{X}_0 | \mathcal{F}_{-k}^k)$ we have for every $\varepsilon > 0$:

$$\begin{aligned} P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1 > \varepsilon) & \leq P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_1^p > \varepsilon^p) \\ & \leq P(|\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_p^p > \varepsilon^p) \leq \frac{1}{\varepsilon^p} E |\mathbf{X}_0 - f_k(\mathbf{Z}_{-k}, \dots, \mathbf{Z}_k)|_p^p \leq \frac{a_{p,k}^p}{\varepsilon^p}. \end{aligned}$$

By choosing $\Phi(\varepsilon) = \varepsilon^{-p}$ and $a_k = a_{p,k}^p$ we have $a_k \rightarrow 0$ as $k \rightarrow \infty$, and $(\mathbf{X}_n)_{n \in \mathbb{Z}}$ is hence P -NED on $(\mathbf{Z}_n)_{n \in \mathbb{Z}}$. \square

Towards the proof of Theorem 2.6 we state the following lemma.

Lemma A.1. *Let $((X_i, Y_i))_{i \geq 1}$ be a stationary process with Lipschitz continuous marginal distribution function F . The kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by (1) satisfies the variation condition.*

Proof. Let (X, Y) and (X', Y') be independent copies of (X_1, Y_1) . Since h is an indicator function we have

$$E \left[\sup_{|(x,y,x',y') - (X,Y,X',Y')|_2 \leq \varepsilon} |h((x,y), (x',y')) - h((X,Y), (X',Y'))| \right] = P(\Omega_0),$$

where Ω_0 is the set of all $\omega \in \Omega$ for which there is a point (x, y, x', y') in the ε -ball around $(X(\omega), Y(\omega), X'(\omega), Y'(\omega)) \in \mathbb{R}^4$ such that exactly one of $(X(\omega) - X'(\omega))(Y(\omega) - Y'(\omega))$ and $(x - x')(y - y')$ is positive. Then $\omega \in \Omega_0$ implies $(X(\omega) - X'(\omega), Y(\omega) - Y'(\omega)) \in A_0$ with

$$A_0 = \left\{ (s, t) \in \mathbb{R}^2 \mid \min(|s|, |t|) \leq \sqrt{2}\varepsilon \right\}.$$

Letting $\varepsilon' = \sqrt{2}\varepsilon$,

$$\begin{aligned} P(\Omega_0) & \leq P((X - X', Y - Y') \in A_0) \leq P(|X - X'| \leq \varepsilon') + P(|Y - Y'| \leq \varepsilon') \\ & = \int_{\mathbb{R}} (F_X(t + \varepsilon') - F_X(t - \varepsilon')) dF_X(t) + \int_{\mathbb{R}} (F_Y(t + \varepsilon') - F_Y(t - \varepsilon')) dF_Y(t) \leq 4L_0 \varepsilon' \end{aligned}$$

where L_0 is the Lipschitz constant of F . Hence the variation condition is fulfilled with $L = 4\sqrt{2}L_0$. \square

Proof of Theorem 2.6 (Asymptotic distribution of \hat{T}_n). We use the U -statistic representation of Kendall's tau,

$$\hat{\tau}_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} h((x_1, y_1), (x_2, y_2))$$

where the kernel function $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by (1). The kernel h is bounded and satisfies the variation condition (Lemma A.1). By first applying a bounded, strictly increasing transformation to both margin processes $(X_i)_{i \geq 1}$ and $(Y_i)_{i \geq 1}$, which leaves Kendall's tau and hence $\sqrt{n}\lambda(\tau_{[n\lambda]} - \hat{\tau}_n)_{0 \leq \lambda \leq 1}$ unchanged, we conclude from Corollary 3.4 that

$$\sqrt{n}\lambda(\tau_{[n\lambda]} - \hat{\tau}_n)_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} 2\sigma_h(B(\lambda))_{0 \leq \lambda \leq 1},$$

where

$$(26) \quad \sigma_h^2 = \text{Var}(h_1(X_1, Y_1)) + 2 \sum_{i=2}^{\infty} \text{Cov}(h_1(X_1, Y_1), h_1(X_i, Y_i)).$$

Here h_1 denotes the linear part of the kernel h in the Hoeffding decomposition, cf. (11). To convince ourselves that σ_h^2 coincides with D^2 given in (6), we need to check that $h_1 = \psi$. First note that $\tau = E(h((X, Y), (X', Y')))$, where (X, Y) and (X', Y') are two independent copies drawn from F . The first order term of the Hoeffding decomposition is then given by

$$\begin{aligned} h_1(x, y) &= E(h((x, y), (X, Y))) - \tau = P((X - x)(Y - y) > 0) - \tau \\ &= P(X < x, Y < y) + P(X > x, Y > y) - \tau. \end{aligned}$$

Since the one-dimensional marginals $F_X(x) = P(X \leq x)$ and $F_Y(y) = P(Y \leq y)$ are continuous, we may express $h_1(x, y)$ as follows

$$\begin{aligned} h_1(x, y) &= P(X \leq x, Y \leq y) + (1 - P(X \leq x) - P(Y \leq y) + P(X \leq x, Y \leq y)) - \tau \\ &= 2F(x, y) - F_X(x) - F_Y(y) + 1 - \tau = \psi(x, y). \end{aligned}$$

Finally, by the continuous mapping theorem, applied to $f \mapsto \sup |f|$, we have

$$\sqrt{n} \sup_{0 \leq \lambda \leq 1} \lambda |\tau_{[n\lambda]} - \hat{\tau}_n| \xrightarrow{\mathcal{D}} 2D \sup_{0 \leq \lambda \leq 1} |B(\lambda)|,$$

and by noting that $\sup_{0 \leq \lambda \leq 1} \lambda |\tau_{[n\lambda]} - \hat{\tau}_n| - \sup_{0 \leq k \leq n} \frac{k}{n} |\hat{\tau}_k - \hat{\tau}_n| = o_P(1/\sqrt{n})$, we have proved (5). \square

Proof of Theorem 2.9 (Consistency of the variance estimator). With ψ being the linear part of the kernel h , cf. (1), we have $E\psi(X_1, Y_1) = 0$. Letting

$$\tilde{X}_{n,t} = n^{-1/2}\psi(X_t, Y_t),$$

we know from (30) with ψ in the role of g_1 that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{t=1}^n \sum_{s=1}^n E(\tilde{X}_{n,t} \tilde{X}_{n,s}) \\ &= \text{Var}(\psi(X_1, Y_1)) + \lim_{n \rightarrow \infty} 2 \sum_{j=2}^n \frac{n-j+1}{n} \text{Cov}(\psi(X_1, Y_1), \psi(X_j, Y_j)) = D^2. \end{aligned}$$

Applying Theorem 2.1 of de Jong and Davidson (2000) to the array $(\tilde{X}_{n,t})_{n \in \mathbb{N}, t=1, \dots, n}$ yields

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \kappa \left(\frac{t-s}{b_n} \right) \psi(X_t, Y_t) \psi(X_s, Y_s) \\ &= \frac{1}{n} \sum_{i=1}^n \psi(X_i, Y_i)^2 + \frac{2}{n} \sum_{j=1}^{n-1} \kappa \left(\frac{j}{b_n} \right) \sum_{i=1}^{n-j} \psi(X_i, Y_i) \psi(X_{i+j}, Y_{i+j}) \xrightarrow{p} D^2. \end{aligned}$$

Parts (i) and (ii) of Lemma 2.5 together ensure that the assumptions of the theorem are met. It remains to show that the convergence also holds if $\psi(X_i, Y_i)$ is replaced by $\hat{\psi}_{n,i}$, i.e.,

$$(27) \quad \frac{1}{n} \sum_{i=1}^n \hat{\psi}_{n,i}^2 + \frac{2}{n} \sum_{j=1}^{n-1} \kappa \left(\frac{j}{b_n} \right) \sum_{i=1}^{n-j} \hat{\psi}_{n,i} \hat{\psi}_{n,i+j} \xrightarrow{p} D^2.$$

In order to prove (27) we abbreviate $\psi(X_i, Y_i)$ by ψ_i and show

$$\begin{aligned} \text{(I)} \quad & \frac{1}{n} \sum_{i=1}^n (\hat{\psi}_{n,i}^2 - \psi_i^2) \xrightarrow{p} 0, \\ \text{(II)} \quad & \frac{1}{n} \sum_{j=1}^{n-1} \kappa \left(\frac{j}{b_n} \right) \sum_{i=1}^{n-j} (\hat{\psi}_{n,i} \hat{\psi}_{n,i+j} - \psi_i \psi_{i+j}) \xrightarrow{p} 0. \end{aligned}$$

The main tool for proving both statements (I) and (II) is Theorem 3.5. Recall

$$\begin{aligned} \psi_i &= 2F(X_i, Y_i) - F_X(X_i) - F_Y(Y_i) + 1 - \tau, \\ \hat{\psi}_{n,i} &= 2\mathbb{F}_n(X_i, Y_i) - \mathbb{F}_{X,n}(X_i) - \mathbb{F}_{Y,n}(Y_i) + 1 - \hat{\tau}_n. \end{aligned}$$

Theorem 3.5 states that the empirical process

$$(\sqrt{n}(\mathbb{F}_n(x, y) - F(x, y)))_{x, y \in \mathbb{R}}$$

converges weakly to a two-dimensional Gaussian limit process in $D(\mathbb{R}^2)$. By the continuous mapping theorem, applied to $f \mapsto \sup |f|$, we have that

$$(28) \quad S_n = \sup_{x, y \in \mathbb{R}} \sqrt{n} |\mathbb{F}_n(x, y) - F(x, y)|$$

also converges in distribution and is hence stochastically bounded. As a consequence of Theorem 3.2, it also holds $\sqrt{n}(\hat{\tau}_n - \tau) = O_P(1)$ for $n \rightarrow \infty$. Furthermore we note that

$$(29) \quad \sqrt{n} |\mathbb{F}_{X,n}(x) - F_X(x)| = \lim_{y \rightarrow \infty} \sqrt{n} |\mathbb{F}_n(x, y) - F(x, y)| \leq S_n,$$

likewise for $\mathbb{F}_{Y,n}$ and $F_Y(x)$, and $|\hat{\psi}_{n,i} + \psi_i| \leq 6$ for all i and n . Analyzing the expression in (I) we find

$$\begin{aligned} & \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\psi}_{n,i}^2 - \psi_i^2) \right| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n |\hat{\psi}_{n,i} + \psi_i| |\hat{\psi}_{n,i} - \psi_i| \leq \frac{6}{\sqrt{n}} \sum_{i=1}^n |\hat{\psi}_{n,i} - \psi_i| \\ & \leq \frac{6}{n} \sum_{i=1}^n \left[2\sqrt{n} |\mathbb{F}_n(X_i, Y_i) - F(X_i, Y_i)| + \sqrt{n} |\mathbb{F}_{X,n}(X_i) - F_X(X_i)| \right. \\ & \quad \left. + \sqrt{n} |\mathbb{F}_{Y,n}(Y_i) - F_Y(Y_i)| + \sqrt{n} |\hat{\tau}_n - \tau| \right] \\ & \leq \frac{6}{n} \sum_{i=1}^n (4S_n + \sqrt{n} |\hat{\tau}_n - \tau|) = 6(4S_n + \sqrt{n} |\hat{\tau}_n - \tau|) = O_P(1). \end{aligned}$$

Hence

$$\frac{1}{n} \sum_{i=1}^n (\hat{\psi}_{n,i}^2 - \psi_i^2) = O_P(n^{-1/2})$$

and (I) is proved. For (II) we first note that by a very similar argumentation as above

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n-j} (\hat{\psi}_{n,i} \hat{\psi}_{n,i+j} - \psi_i \psi_{i+j}) \right| \leq 6(4S_n + \sqrt{n}|\hat{\tau}_n - \tau|) = O_P(1).$$

The second step is then

$$\begin{aligned} & \left| \frac{1}{n} \sum_{j=1}^{n-1} \kappa\left(\frac{j}{b_n}\right) \sum_{i=1}^{n-j} (\hat{\psi}_{n,i} \hat{\psi}_{n,i+j} - \psi_i \psi_{i+j}) \right| \\ & \leq \frac{b_n}{\sqrt{n}} \cdot \left| \frac{1}{b_n} \sum_{j=1}^{n-1} \kappa\left(\frac{j}{b_n}\right) \right| \cdot (24S_n + 6\sqrt{n}|\hat{\tau}_n - \tau|) \end{aligned}$$

The first factor b_n/\sqrt{n} converges to zero, the second factor converges to $\int_0^\infty \kappa(x)dx$ (which is ensured by the existence of an integrable, monotonic dominating function of κ) and is hence bounded. The third factor is stochastically bounded by above's considerations, and hence the product converges to zero in probability. Thus we have proved (II) and consequently (27). Finally, by Slutsky's lemma

$$\frac{\hat{T}_n}{2\hat{D}_n} \xrightarrow{\mathcal{D}} \sup_{0 \leq \lambda \leq 1} |B(\lambda)|.$$

The proof is complete. \square

APPENDIX B. PROOFS OF SECTION 3

Proof of Theorem 3.2 (Invariance principle for U-statistics). From the Hoeffding decomposition we obtain

$$U_{[n\lambda]} = \theta + \frac{2}{[n\lambda]} \sum_{i=1}^{[n\lambda]} g_1(\mathbf{X}_i) + \frac{1}{\binom{[n\lambda]}{2}} \sum_{1 \leq i < j \leq [n\lambda]} g_2(\mathbf{X}_i, \mathbf{X}_j),$$

and hence

$$\sqrt{n} \lambda (U_{[n\lambda]} - \theta) = \frac{2\sqrt{n}\lambda}{[n\lambda]} \sum_{i=1}^{[n\lambda]} g_1(\mathbf{X}_i) + \frac{\sqrt{n}\lambda}{\binom{[n\lambda]}{2}} \sum_{1 \leq i < j \leq [n\lambda]} g_2(\mathbf{X}_i, \mathbf{X}_j).$$

Thus the proof splits into two parts. We will show that of the two terms on the right hand side, the first one is the dominating term that converges to a Brownian motion, while the second term converges to zero uniformly in λ .

Part (1): By parts (i) and (ii) of Lemma 2.5, $(g_1(\mathbf{X}_i))_{i \in \mathbb{Z}}$ is L_2 -NED on $(\mathbf{Z}_i)_{i \in \mathbb{Z}}$ with approximating constants $a_{2,k} = O(k^{-(3+\varepsilon)/2})$. The process $(\mathbf{Z}_i)_{i \in \mathbb{Z}}$ is assumed to be absolutely regular with coefficients $\beta_k = O(k^{-(1+\varepsilon)})$, hence it is also α -mixing with coefficients that decline to zero with at least the same rate. Applying Corollary 3.2 of Wooldridge and White (1988) (with $g_1(\mathbf{X}_i)$ and \mathbf{Z}_i in the role of Z_i and Y_i , respectively) yields

$$\left(\frac{1}{\sigma_n \sqrt{n}} \sum_{i=1}^{[n\lambda]} g_1(\mathbf{X}_i) \right)_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} (W(\lambda))_{0 \leq \lambda \leq 1},$$

where W is a standard Brownian motion and

$$(30) \quad \sigma_n^2 = \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n g_1(\mathbf{X}_k) \right) = \text{Var}(g_1(\mathbf{X}_1)) + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \text{Cov}(g_1(\mathbf{X}_1), g_1(\mathbf{X}_{k+1})).$$

Under the conditions of Theorem 3.2, the limit variance

$$\sigma^2 = \text{Var}(g_1(\mathbf{X}_1)) + 2 \sum_{i=2}^{\infty} \text{Cov}(g_1(\mathbf{X}_1), g_1(\mathbf{X}_i))$$

is finite, cf. Theorem 2.3 of Ibragimov (1962), and σ_n^2 converges to σ^2 as $n \rightarrow \infty$. Thus, by Slutsky's lemma we also have

$$\left(\frac{1}{\sqrt{n}} \left(\frac{\lambda n}{[\lambda n]} \right) \sum_{i=1}^{[n\lambda]} g_1(\mathbf{X}_i) \right)_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} (\sigma W(\lambda))_{0 \leq \lambda \leq 1}.$$

Part (2): Since

$$\sup_{0 \leq \lambda \leq 1} \frac{\sqrt{n} \lambda}{\binom{[n\lambda]}{2}} \left| \sum_{1 \leq i < j \leq [n\lambda]} g_2(\mathbf{X}_i, \mathbf{X}_j) \right| \leq \sup_{1 \leq k \leq n} \frac{k+1}{\sqrt{n}} |U_k(g_2)|,$$

we will show that the right-hand side converges to zero almost surely. By Lemma 2.5 (ii) and our assumption that $(\mathbf{X}_i)_{i \in \mathbb{Z}}$ is bounded, we have that $(\mathbf{X}_i)_{i \in \mathbb{Z}}$ is L_1 -NED on $(\mathbf{Z}_i)_{i \in \mathbb{Z}}$ with approximating constants $a_{1,k} = O(k^{-(3+\varepsilon)})$, and further

$$(31) \quad \sum_{k=1}^n k \left(\beta_k + \left(\sum_{i=k}^{\infty} a_{1,i} \right)^{\frac{1}{2}} \right) = O(n^{1-\delta})$$

for any $0 < \delta \leq \min(1, \varepsilon/2)$. Dehling and Wendler (2010, Theorem 1) show that under condition (31)

$$\frac{k^{1/2 + \delta/2}}{\log^{3/2} k \log \log k} U_k(g_2) \xrightarrow{a.s.} 0.$$

Here we make use of the fact that along with the kernel g also its degenerate part g_2 fulfills the variation condition. Dehling and Wendler (2010) consider U -statistics of one-dimensional processes $(X_k)_{k \geq 1}$, but the results holds true for U -statistics of multivariate processes as well. Hence

$$\frac{k+1}{\sqrt{k}} U_k(g_2) \xrightarrow{a.s.} 0 \quad (k \rightarrow \infty)$$

and further

$$\frac{1}{\sqrt{n}} \sup_{1 \leq k \leq n} (k+1) U_k(g_2) \xrightarrow{a.s.} 0.$$

as $n \rightarrow \infty$. The proof is complete. \square

Proof of Corollary 3.4. We consider the functional on $D[0, 1]$ given by

$$x(t) \mapsto x(t) - t x(1), \quad 0 \leq t \leq 1.$$

This is a continuous functional, which, applied to the process on the left hand side of (12), yields

$$\sqrt{n} \lambda (U_{[n\lambda]} - \theta) - \lambda \sqrt{n} (U_n - \theta) = \sqrt{n} \lambda (U_{[n\theta]} - U_n).$$

Thus we may apply the continuous mapping theorem to obtain (14). \square

Proof of Theorem 3.5 (Empirical process invariance principle). Without loss of generality, we can assume that X_n and Y_n have a uniform distribution on the interval $[0, 1]$, otherwise we consider the random variables $F_X(X_n)$ and $F_Y(Y_n)$. By Lemma 2.5 (i), the P -NED condition still holds after this transformation. By Lemma 2.5 (i) and Proposition 2.11 of Borovkova, Burton, and Dehling (2001), have that for any s, t , the sequence

$$(\mathbf{1}_{\{X_n \leq s, Y_n \leq t\}})_{n \in \mathbb{N}}$$

is L_1 -NED with approximation constants $a_{1,k} = O(k^{-6-\varepsilon'})$ for some $\varepsilon' > 0$. The finite-dimensional convergence of the process

$$(W_n(s, t))_{s, t \in \mathbb{R}} = (\sqrt{n}(F_n(s, t) - F(s, t)))_{s, t \in \mathbb{R}}$$

follows from Theorem 4 of Borovkova et al. (2001) together with the Cramér-Wold device. For the tightness of the process, we have to show that for every $\varepsilon, \delta > 0$, there is a $k \in \mathbb{N}$, such that

$$(32) \quad \limsup_{n \rightarrow \infty} P\left(\max_{k_1, k_2=1, \dots, k} \sup_{\substack{\frac{k_1-1}{k} \leq s \leq \frac{k_1}{k} \\ \frac{k_2-1}{k} \leq t \leq \frac{k_2}{k}}} |W_n(s, t) - W_n(\frac{k_1}{k}, \frac{k_2}{k})| > \delta\right) \leq \varepsilon.$$

By the Lipschitz continuity of F , we have that $|F(s, t) - F(s', t')| \leq |s - s'| + |t - t'|$, so by the monotonicity of F and F_n , we can conclude that for $s \leq s'' \leq s', t \leq t'' \leq t'$

$$|W_n(s'', t'') - W_n(s, t)| \leq |W_n(s', t') - W_n(s, t)| + \sqrt{n}(|s - s'| + |t - t'|).$$

Let $k = 2^l$ for an $l \in \mathbb{N}$ to be chosen later. Let $\tilde{l} \in \mathbb{N}$, such that $\tilde{l} > l$ and $\frac{\delta}{8\sqrt{n}} \leq 2^{-\tilde{l}} \leq \frac{\delta}{4\sqrt{n}}$. Instead of (32), it suffices to show that

$$\limsup_{n \rightarrow \infty} P\left(\max_{k_1, k_2=1, \dots, k} \max_{i, j=1, \dots, 2^{l-\tilde{l}}} |W_n(\frac{k_1}{k} + \frac{i}{2^{\tilde{l}}}, \frac{k_2}{k} + \frac{j}{2^{\tilde{l}}}) - W_n(\frac{k_1}{k}, \frac{k_2}{k})| > \frac{\delta}{2}\right) \leq \varepsilon.$$

For any two intervals $I_1, I_2 \subset \mathbb{R}$ define

$$Q_n(I_1 \times I_2) = \frac{1}{n} \left(\sum_{i=1}^n \mathbf{1}_{\{(X_i, Y_i) \in I_1 \times I_2\}} - P((X_1, Y_1) \in I_1 \times I_2) \right).$$

Similar to Lemma A.1, one can prove that the indicator of $I_1 \times I_2$ satisfies the variation condition (Definition 3.1). So we can apply Lemma 3.1 of Borovkova et al. (2001) to obtain

$$EQ_n^4(I_1 \times I_2) \leq C \left(P^{1+\delta'}((X_1, Y_1) \in I_1 \times I_2) + n^{-(1+\delta')} \right)$$

for some $\delta' > 0$. We now use a chaining technique (see for example van der Vaart and Wellner (1996), proof of Theorem 2.2.4, for more details) and get

$$\begin{aligned} & \left(E \left(\max_{k_1, k_2=1, \dots, k} \max_{i, j=1, \dots, 2^{l-\tilde{l}}} |W_n(\frac{k_1}{k} + \frac{i}{2^{\tilde{l}}}, \frac{k_2}{k} + \frac{j}{2^{\tilde{l}}}) - W_n(\frac{k_1}{k}, \frac{k_2}{k})| \right)^4 \right)^{1/4} \\ & \leq \left(E \left(\sum_{l_1, l_2} \max_{\substack{i=1, \dots, 2^{l_1} \\ j=1, \dots, 2^{l_2}}} \left| Q_n \left(\left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) \right| \right)^4 \right)^{1/4} \end{aligned}$$

$$\begin{aligned}
& \leq \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\sum_{\substack{i=1, \dots, 2^{l_1} \\ j=1, \dots, 2^{l_2}}} E \left(Q_n \left(\left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) \right)^4 \right)^{\frac{1}{4}} \\
& \leq \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\sum_{\substack{i=1, \dots, 2^{l_1} \\ j=1, \dots, 2^{l_2}}} C \left(P^{1+\delta'} \left((X_1, X_2) \in \left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) + n^{-(1+\delta')} \right) \right)^{1/4} \\
(33) \quad & \leq C \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\sum_{\substack{i=1, \dots, 2^{l_1} \\ j=1, \dots, 2^{l_2}}} P^{1+\delta'} \left((X_1, X_2) \in \left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) \right)^{1/4} \\
(34) \quad & + C \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\sum_{\substack{i=1, \dots, 2^{l_1} \\ j=1, \dots, 2^{l_2}}} n^{-(1+\delta')} \right)^{1/4}
\end{aligned}$$

We will treat the two summands (33) and (34) of the last line separately. By the Lipschitz continuity of F , we have that

$$\begin{aligned}
& P \left((X_1, X_2) \in \left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) \\
& \leq \min \left\{ P \left((X_1, X_2) \in \left[\frac{i-1}{2^{l_1}}, \frac{i}{2^{l_1}} \right] \times \mathbb{R} \right), P \left((X_1, X_2) \in \mathbb{R} \times \left[\frac{j-1}{2^{l_2}}, \frac{j}{2^{l_2}} \right] \right) \right\} \\
& = \min \left\{ \frac{1}{2^{l_1}}, \frac{1}{2^{l_2}} \right\} = \frac{1}{\max\{2^{l_1}, 2^{l_2}\}}.
\end{aligned}$$

With $\delta'' = \delta'/4$, we find that the first summand (33) is bounded from above by

$$C \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\frac{1}{\max\{2^{l_1}, 2^{l_2}\}} \right)^{\delta''} \leq \sum_{l'=l}^{\infty} l' 2^{-\delta'' l'}.$$

As the last series is summable, we can make it arbitrary small by choosing l , and thus $k = 2^l$, large enough. Since $2^{\tilde{l}} \leq 8\delta^{-1}\sqrt{n}$ and therefore $\tilde{l} \leq C \log n$, we have that the second summand (34) is bounded by

$$C \sum_{\substack{l_1, l_2 \\ l \leq \max\{l_1, l_2\} \leq \tilde{l}}} \left(\left(\frac{8\sqrt{n}}{\delta} \right)^2 n^{-(1+\delta')} \right)^{1/4} \leq C \log^2(n) n^{-\frac{\delta'}{4}},$$

which converges to zero as $n \rightarrow \infty$.

□

APPENDIX C. PROOFS OF SECTION 4

Proof of Theorem 4.3 (Change-point identification). The proof relies on the fact that

$$(35) \quad (C_n(\lambda))_{0 \leq \lambda \leq 1} = \left(\frac{[\lambda n]}{n} (\hat{\tau}_{[\lambda n]} - \hat{\tau}_n) \right)_{0 \leq \lambda \leq 1} \xrightarrow{\mathcal{D}} (c(\lambda))_{0 \leq \lambda \leq 1}$$

in $D[0, 1]$, where the deterministic function c is defined in (15). With the argmax theorem (van der Vaart and Wellner, 1996, Corollary 3.2.3) we have that under Assumption 4.2

$$\hat{b}_n = \arg \max_{0 \leq \lambda \leq 1} |C_n(\lambda)| \xrightarrow{p} \arg \max_{0 \leq \lambda \leq 1} |C(\lambda)| = b.$$

It remains to prove (35). To simplify notation, we let $m = [bn]$, write $\mathbf{Z}_{i,n}$ short for $(X_{i,n}, Y_{i,n})$ and further suppress the subscript n . Assume for an instant that the \mathbf{Z}_i , $i = 1, \dots, n$, are independent. Then we have

$$t_n(k) = E(\hat{\tau}_k) = \begin{cases} \tau_F & \text{for } k \leq m, \\ \frac{m(m-1)}{k(k-1)}\tau_F + \frac{(k-m)(k-m-1)}{k(k-1)}\tau_G + \frac{2m(k-m)}{k(k-1)}\tau_{FG} & \text{for } k \geq m+1, \end{cases}$$

from where we derive the mean function

$$c_n(\lambda) = E[C_n(\lambda)] = \frac{[\lambda n]}{n} (t_n([\lambda n]) - t_n(n)), \quad 0 \leq \lambda \leq 1,$$

of the process of $(C_n(\lambda))_{0 \leq \lambda \leq 1}$ and observe that it converges to c . Thus it remains to show that

$$(36) \quad \max_{0 \leq \lambda \leq 1} |C_n(\lambda) - c_n(\lambda)| \xrightarrow{p} 0$$

also under the short-range dependence assumption of Model 4.1. Although the \mathbf{Z}_i , $i = 1, \dots, n$ are weakly dependent in the following, c_n and $t_n(k)$ are still defined as above, assuming independent observations. In what follows, we will prove that

$$(37) \quad \max_{m < k \leq n} \frac{k}{n} |\hat{\tau}_k - t_n(k)| \xrightarrow{p} 0.$$

The convergence of $\max_{1 < k \leq m} \frac{k}{n} |\hat{\tau}_k - C(\frac{k}{n})|$ follows along the same lines. Hence the maximum in (37) can be extended to the range $1 \leq k \leq n$ and (36) follows.

We split the difference $\frac{k}{n} |\hat{\tau}_k - t_n(k)|$ into three parts: two one-sample U -statistics and one two-sample U -statistic. By the triangle inequality we get:

$$(38) \quad \frac{k}{n} |\hat{\tau}_k - t_n(k)| \leq \left| \frac{2}{n(k-1)} \sum_{1 \leq i < j \leq m} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_F) \right|$$

$$(39) \quad + \left| \frac{2}{n(k-1)} \sum_{m+1 \leq i < j \leq k} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_G) \right|$$

$$(40) \quad + \left| \frac{2}{n(k-1)} \sum_{1 \leq i \leq m < j \leq k} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_{FG}) \right|.$$

For the first summand (38), we have by Theorem 3.2:

$$\max_{m < k \leq n} \left| \frac{2}{n(k-1)} \sum_{1 \leq i < j \leq m} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_F) \right| \leq \left| \frac{2}{n(m-1)} \sum_{1 \leq i < j \leq m} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_F) \right| \xrightarrow{p} 0.$$

Due to our assumptions, $(\mathbf{Z}_{i,n})_{m+1 \leq i \leq n}$ is a stationary process which satisfies condition (9), so we also treat the second summand (39) by Theorem 3.2:

$$\max_{m < k \leq n} \left| \frac{2}{n(k-1)} \sum_{m+1 \leq i < j \leq k} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_G) \right| \xrightarrow{p} 0.$$

For the third summand (40), we apply a two-sample Hoeffding decomposition

$$h(\mathbf{z}_1, \mathbf{z}_2) = \tau_{FG} + \tilde{h}_1(\mathbf{z}_1) + \tilde{h}_2(\mathbf{z}_2) + \tilde{h}_3(\mathbf{z}_1, \mathbf{z}_2)$$

with

$$\tilde{h}_1(\mathbf{z}_1) = E(h(\mathbf{z}_1, \mathbf{Z}_{m+1})) - \tau_{FG},$$

$$\tilde{h}_2(\mathbf{z}_2) = E(h(\mathbf{Z}_1, \mathbf{z}_2)) - \tau_{FG},$$

$$\tilde{h}_3(\mathbf{z}_1, \mathbf{z}_2) = h(\mathbf{z}_1, \mathbf{z}_2) - \tilde{h}_1(\mathbf{z}_1) - \tilde{h}_2(\mathbf{z}_2) - \tau_{FG},$$

where $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^2$. We get

$$\begin{aligned} & \frac{2}{n(k-1)} \sum_{1 \leq i \leq m < j \leq k} (h(\mathbf{Z}_i, \mathbf{Z}_j) - \tau_{FG}) \\ (41) \quad &= \frac{2}{n} \sum_{i=1}^m \tilde{h}_1(\mathbf{Z}_i) + \frac{2}{n} \sum_{j=m+1}^k \tilde{h}_2(\mathbf{Z}_j) + \frac{2}{n(k-1)} \sum_{1 \leq i \leq m < j \leq k} \tilde{h}_3(\mathbf{Z}_i, \mathbf{Z}_j). \end{aligned}$$

To obtain a maximal inequality, we use Theorem 2.4.1 of Stout (1974): For random variables R_1, \dots, R_n with $E\left(\sum_{j=k}^{k+l-1} R_j\right) \leq C_1 l$ for a constant C_1 , we have that

$$(42) \quad E \max_{1 \leq l \leq n} \left(\sum_{j=1}^l R_j \right)^2 \leq C_1 n \left(\frac{\log(2n)}{\log 2} \right)^2.$$

We define the random variables $R_j = \sum_{i=1}^m \tilde{h}_3(\mathbf{Z}_i, \mathbf{Z}_{j+m})$. Without loss of generality, we can assume that the random variables \mathbf{Z}_i are bounded and thus the process is L_1 -NED by Lemma 2.5. Furthermore, the kernel \tilde{h}_3 is degenerate, so we can apply Proposition 6.2 of Dehling and Fried (2012) to obtain the moment bound

$$E \left(\sum_{j=1}^{k+l-1} R_j \right)^2 = E \left(\sum_{j=1}^{k+l-1} \sum_{i=1}^m \tilde{h}_3(\mathbf{Z}_i, \mathbf{Z}_{j+m}) \right)^2 \leq C m l.$$

Applying (42) it follows that

$$\begin{aligned} & E \left(\max_{m < k \leq n} \left| \frac{2}{n(k-1)} \sum_{1 \leq i \leq m < j \leq k} \tilde{h}_3(\mathbf{Z}_i, \mathbf{Z}_j) \right| \right)^2 \\ & \leq \frac{1}{m^2} E \left(\max_{1 < k \leq n-m} \left| \frac{2}{n} \sum_{j=1}^k R_j \right| \right)^2 \leq C \frac{\log^2(2n)}{mn \log^2 2} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. Thus the third summand in (41) converges to zero in probability. As for the first two summands, we have that $E \left(\sum_{j=k}^{k+l-1} \tilde{h}_2(\mathbf{Z}_j) \right)^2 \leq C l$, since

$$\frac{1}{l} \text{Var} \left(\sum_{j=k}^{k+l-1} \tilde{h}_2(\mathbf{Z}_j) \right)$$

converges to a finite limit as $l \rightarrow \infty$. Hence, (42) applied to $R_j = \tilde{h}_2(\mathbf{Z}_{j+m})$ leads to

$$\max_{m < k \leq n} \left| \frac{2}{n} \sum_{j=m+1}^k \tilde{h}_2(\mathbf{Z}_j) \right| \xrightarrow{p} 0$$

as $n \rightarrow \infty$. Finally, $\frac{2}{n} \sum_{i=1}^m \tilde{h}_1(\mathbf{Z}_i) \xrightarrow{p} 0$, as its variance converges to zero. We have thus shown (37), which completes the proof. \square

APPENDIX D. MISC

Lemma D.1. *Let (X_1, Y_1, X_2, Y_2) be jointly Gaussian with covariance matrix*

$$S = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ \varrho & 0 & 1 & \\ 0 & \varrho & 0 & 1 \end{pmatrix},$$

i.e., (X_1, X_2) and (Y_1, Y_2) are independent, and each has a bivariate normal distribution with correlation ϱ . Then

$$\text{Cov}(\psi(X_1, Y_1), \psi(X_2, Y_2)) = \frac{1}{\pi^2} \arcsin^2\left(\frac{\varrho}{2}\right),$$

where ψ is as in Theorem 2.6.

Proof. Letting $U_i = F_X(X_i)$, $i = 1, 2$, (i.e., (U_1, U_2) have the Gauss copula with correlation ϱ as joint distribution function), straightforward calculus yields

$$(43) \quad \text{Cov}(\psi(X_1, Y_1), \psi(X_2, Y_2)) = 2E(U_1 U_2) (2E(U_1 U_2) - 1) + \frac{1}{4}.$$

Croux and Dehon (2010) give the following expression for the population version s of Spearman's rho at a bivariate normal distribution with correlation ϱ :

$$s = \frac{6}{\pi} \arcsin\left(\frac{\varrho}{2}\right),$$

a result which can be traced back to Pearson (1907). Together with (22) we deduce

$$(44) \quad E(U_1 U_2) = \frac{1}{2\pi} \arcsin\left(\frac{\varrho}{2}\right) + \frac{1}{4}.$$

Plugging (44) into (43) yields the stated result. \square

REFERENCES

- A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 2009.
- I. Berkes, S. Hörmann, and J. Schauer. Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.*, 119:1298–1324, 2009.
- H. J. Bierens. *Robust Methods and Asymptotic Theory*. Lecture Notes in Economics and Mathematical Statistics 192. Berlin: Springer-Verlag, 1981.
- M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. Springer Texts in Statistics. New York: Springer, 1999.
- S. Borovkova, R. M. Burton, and H. Dehling. Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Trans. Amer. Math. Soc.*, 353(11):4261–4318, 2001.
- R. Campbell, C. S. Forbes, K. Koedijk, and P. Kofman. Diversification meltdown or just fat tails. Technical report, Monash University, Australia, 2003.
- C. Croux and C. Dehon. Influence functions of the Spearman and Kendall correlation measures. *Stat. Methods Appl.*, 19(4):497–515, 2010.

- M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Chichester: J. Wiley & Sons, 1997.
- R. M. de Jong and J. Davidson. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2):407–424, 2000.
- H. Dehling and R. Fried. Asymptotic distribution of two-sample empirical u-quantiles with applications to robust tests for shifts in location. *J. Multivariate Anal.*, 105:124–140, 2012.
- H. Dehling and M. Wendler. Law of the iterated logarithm for U -statistics of weakly dependent observations. In I. Berkes, R. Bradley, H. Dehling, M. Peligrad, and R. Tichy, editors, *Dependence in Probability, Analysis and Number Theory*, pages 177–194. Kendrick Press, 2010.
- B. Dengler. *On the Asymptotic Behaviour of the Estimator of Kendall's Tau*. PhD thesis, Technische Universität Wien, Austria, 2010.
- M. Denker and G. Keller. On U -statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, 64:505–522, 1983.
- M. Denker and G. Keller. Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics*, 44(1/2):67–93, 1986.
- P. Galeano and D. Peña. Covariance changes detection in multivariate time series. *Journal of Statistical Planning and Inference*, 137(1):194–211, 2007.
- I. A. Ibragimov. Some limit theorems for stationary processes. *Theor. Probab. Appl.*, 7:349–382, 1962.
- I. A. Ibragimov and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff, 1971.
- C. Kao, L. Trapani, and G. Urga. Testing for instability in covariance structures. *CPR Working Paper # 131*, Syracuse University, Syracuse, NY, 2011.
- P. A. P. Moran. Rank correlation and product-moment correlation. *Biometrika*, 35:203–206, 1948.
- R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. New York: Springer, 2nd edition, 2006.
- K. Pearson. *Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation*. Drapers' company research memoirs: Biometric series. London: Dulau & Co., 1907.
- B. M. Pötscher and I. R. Prucha. *Dynamic Nonlinear Econometric Models*. Berlin: Springer-Verlag, 1997.
- L. Rüschendorf. Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.*, 4(5):912–923, 1976.
- W. F. Stout. *Almost Sure Convergence*. London: Academic Press, 1974.
- A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer Series in Statistics. New York: Springer-Verlag, 1996.
- M. Wendler. *Empirical U-Quantiles of Dependent Data*. PhD thesis, Ruhr-Universität Bochum, Germany, 2011.
- D. Wied, H. Dehling, M. van Kampen, and D. Vogel. A fluctuation test for constant Spearman's rho. *SFB 823 Discussion Paper 2011/16*, 2011.
- D. Wied, W. Krämer, and H. Dehling. Testing for a change in correlation at an unknown point in time using an extended delta method. *Econometric Theory*, 28(3):570–589, 2012.
- C. S. Withers. Conditions for linear processes to be strong-mixing. *Z. Wahrsch. Verw. Gebiete*, 57:477–480, 1981.
- J. M. Wooldridge and H. White. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4:210–230, 1988.
- K.-I. Yoshihara. Limiting behavior of U -statistics for stationary, absolutely regular processes. *Z. Wahrsch. Verw. Gebiete*, 35:237–252, 1976.

FAKULTÄT FÜR MATHEMATIK, RUHR-UNIVERSITÄT BOCHUM, 44780 BOCHUM, GERMANY
E-mail address: herold.dehling@rub.de, vogeltds@rub.de, martin.wendler@rub.de

FAKULTÄT STATISTIK, TECHNISCHE UNIVERSITÄT DORTMUND, 44221 DORTMUND, GERMANY
E-mail address: wied@statistik.tu-dortmund.de